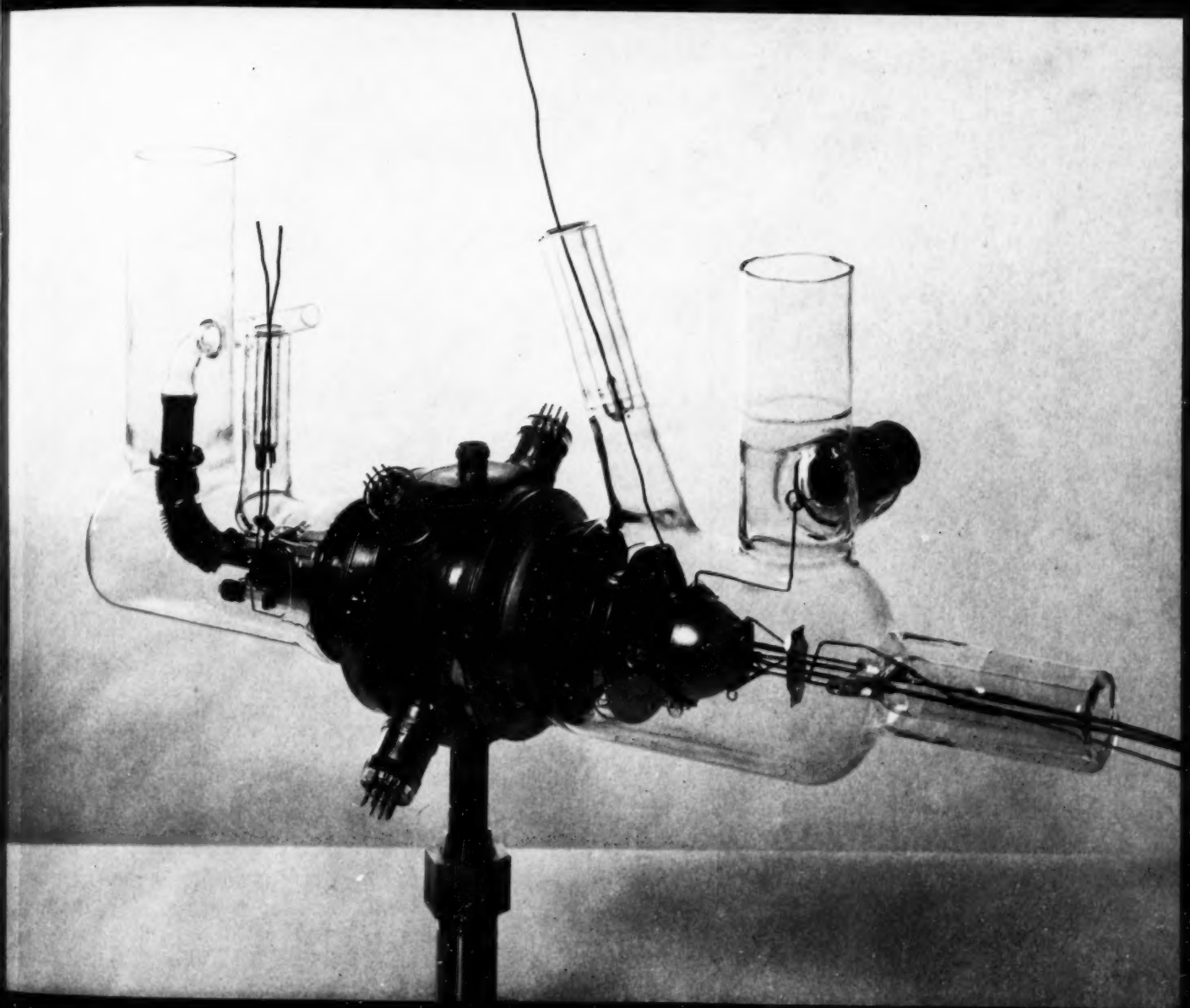
JOURNAL OF APPLIED PHYSICS



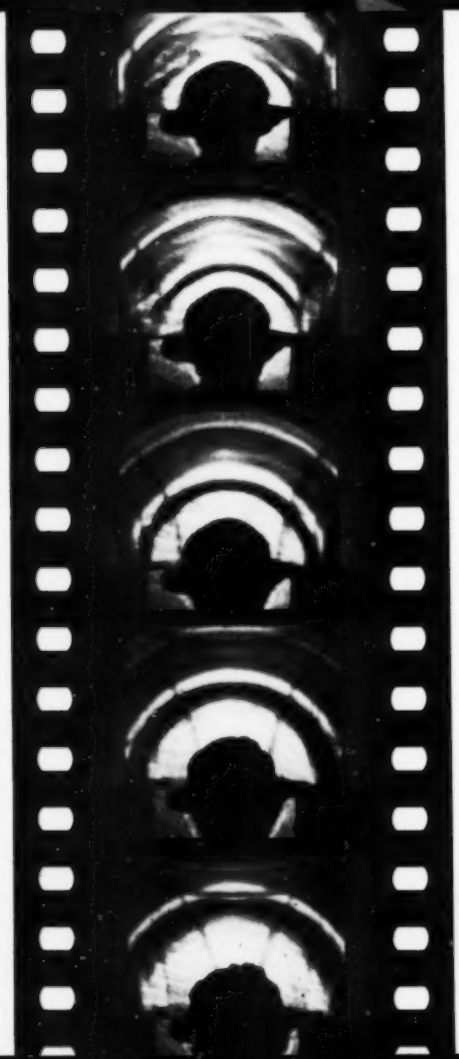
FRAMING CAMERA

magnifying time

The six frames in the strip at left are selected from a sequence of 25 frames taken at 720,000 fps and show shock-wave progression through a block of Lucite.

Exposure: approx. microseconds.

Lighting: schlieren by lighting—exploded wire source



Suited particularly for studies of combustion, corona-discharge, explosion, plastic- and elastic-deformation, and shock-wave events, Beckman & Whitley Framing Cameras are made in models, as shown, capable of up to 2.4 million frames per second, and special models for even higher speeds.

The image formed by the objective lens is transferred from the mirror of a Kinder Turbine through a precision multi-channel optical system to produce 25 frames. With a frame size of 3/4 by 1 inches, the individual pictures have fractional-microsecond exposure times. Optical resolution in the time direction is 50 lines per millimeter while in the space direction it is 100 lines per millimeter. Effective aperture at the film is approximately f/14.5.

In the operating rack, controls are provided for the operator to adjust turbine speed and fire synchronously the event to be recorded. An events-per-unit-time meter gives an accurate measure of the turbine speed.

SEND FOR DETAILS

Beckman & Whitley INC., SAN CARLOS CALIFORNIA

instruments for 4-dimensional documentation

JOURNAL OF APPLIED PHYSICS

Vol. 26, No. 2 FEBRUARY, 1955

In This Issue

Page

Measurement of Microwave Diffraction from a Long Slit in a Thin Conducting Plane . . . . . Jay L. Hirshfield and Clayton M. Zieman 135

Backscattering from Wide-Angle and Narrow-Angle Cones . . . . . Leopold B. Felsen 138

Diffraction by Apertures . . . . . Chaang Huang, Ralph D. Kodis, and Harold Levine 151

Relation of Antimony Transmission and the Photoelectric Yield of Cs-Sb . . . . . Martin Rome 166

Spin-Echo Memory Device . . . . . Sidney Fernbach and Warren G. Proctor 170

Theory of Elasticity and Consolidation for a Porous Anisotropic Solid . . . . . M. A. Biot 182

Radiation of Plasma Noise from Arc Discharge . . . T. Takakura, K. Baba, K. Nunogaki, and H. Mitani 185

Measurement of Carrier Lifetimes in Germanium and Silicon . Donald T. Stevenson and Robert J. Keyes 190

Short-Time Frequency Measurement of Narrow-Band Random Signals by Means of a Zero Counting Process . . . . . Herbert Steinberg, Peter M. Schultheiss, Conrad A. Wogrin, and Felix Zweig 195

Internal Friction of Metal Single Crystals . . . . . Johannes Weertman 202

Anomalous Forward Switching Transient in *p-n* Junction Diodes . . . . . N. T. Jones, R. H. Kingston, and S. F. Neustadter 210

Methods of Measuring Dielectric Constants Based upon a Microwave Network Viewpoint . . . . . A. A. Oliner and H. M. Altschuler 214

Information Theory Aspects of Propagation through Time-Varying Media . . . . . Joseph Feinstein 219

Analog Computer Construction of Conformal Maps in Fluid Dynamics . . . . . N. P. Tomlinson, M. Horowitz, and C. H. Reynolds 229

On the Diffusion of Decaying Particles . . . . . Louis Gold 233

On Power Spectra and the Minimum Detectable Signal in Measurement Systems . . . . . J. J. Freeman 236

Patch Effect for the Thermionic Emission from Polycrystalline Tantalum . . . . . W. B. LaBerge, R. J. Munick, J. A. Dezoteux, J. F. Whalen, and E. A. Coomes 241

Inertial Impaction of Aerosol Particles on Cylinders . . . . J. B. Wong, W. E. Ranz, and H. F. Johnstone 244

Letters to the Editor:

Wall Effects on Microwave Measurements of Ferrite Spheres . E. G. Spencer and R. C. LeCraw 250

Transition Structure of Nickel . . . . . G. I. Finch, K. P. Sinha, and A. Goswami 250

Effect of Superposing a Small Alternating Excitation on the Steady Excitation of a Luminescent Material . . . . . Kenneth F. Stripp and Richard H. Bube 251

Ion Bombardment-Cleaning of Germanium and Titanium as Determined by Low-Energy Electron Diffraction . . . . . H. E. Farnsworth, R. E. Schlier, T. H. George, and R. M. Burger 252

Errata: Relativistic Dynamics of a Charged Particle in Crossed Magnetic and Electric Fields with Application to the Planar Magnetron . . . . . Louis Gold 253

Books Reviewed 254

Books Received 256

Announcements 256

Cover Photograph 256

Index to Advertisers xxii

JOURNAL OF APPLIED PHYSICS

A journal of applications of physics to other sciences and to industry

Published by the

American Institute of Physics

OFFICERS

FREDERICK SEITZ
Chairman

GEORGE B. PEGRAM
Treasurer

GOVERNING BOARD

ROBERT F. BACHER
J. W. BUCHTA
KARL K. DARROW
S. A. GOUDSMIT
DEANE B. JUDD
HUGH S. KNOWLES
C. C. LAURITSEN
W. H. MARKWOOD, JR.
WILLIAM F. MEGGERS
PHILIP M. MORSE
BRIAN O'BRIEN
HARRY F. OLSON
GEORGE B. PEGRAM
ERIC RODGERS
RALPH A. SAWYER
FREDERICK SEITZ
WILLIAM SHOCKLEY
RICHARD M. SUTTON
J. H. VAN VLECK
MARK W. ZEMANSKY

Consultant on Printing

MELVIN LOOS

ADMINISTRATIVE STAFF

HENRY A. BARTON
Director

WALLACE WATERFALL
Executive Secretary

THEODORE VORBURGER
Advertising Manager

RUTH F. BRYANS
Publication Manager

EDITH I. NEFTEL
Circulation Manager

KATHRYN SETZE
Assistant Treasurer

Editor

Robert L. Sproull

Editorial Board

Maurice A. Biot
Walker Bleakney
Wallace J. Eckert
Elmer Hutchisson
D. O. North
Louis A. Pipes
W. R. Smythe
Arthur V. Tobolsky
A. M. Weinberg

THE JOURNAL OF APPLIED PHYSICS is devoted to the application of physics to other sciences, to engineering, and to industry. Previous to 1937, this journal was called PHYSICS.

The JOURNAL OF APPLIED PHYSICS welcomes articles of two distinct kinds: (1) Descriptions of new results in experimental or theoretical physics which are of general interest to applied physicists. The author should explain his contribution in such a way that a physicist or engineer in another field of specialization can appreciate its objectives and essential features. (2) Critical reviews of particular areas of applied physics. Authors contemplating writing such reviews should correspond with the Editor before they begin writing.

The JOURNAL OF APPLIED PHYSICS welcomes letters to the editor which describe in a few words the content of laboratory reports and theses and which tell readers how they may obtain copies of such reports.

Manuscripts should be submitted to the Editor, Rockefeller Hall, Cornell University, Ithaca, New York. Authors should read the "Information for Contributors," facing page 1 of the January issue, before writing articles or letters. The *Style Manual* of the American Institute of Physics is also helpful to writers preparing manuscripts for journals published by the American Institute of Physics. It can be obtained for $1.00 from the American Institute of Physics, 57 East 55th Street, New York 22, New York.

Corrected proofs and all correspondence concerning papers in the process of publication should be addressed to the Publication Manager, American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	14.00

Back Number Prices

Yearly back number rate when complete year is available: $15.00.
Single copies: $1.50 each.

Changes of address should be sent to the American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 26, Number 2 February, 1955

Measurement of Microwave Diffraction from a Long Slit in a Thin Conducting Plane

JAY L. HIRSHFIELD AND CLAYTON M. ZIEMAN
USAF Institute of Technology, Wright-Patterson Air Force Base, Ohio
(Received March 23, 1954)

This paper gives the results of measurements of the fields diffracted from a long narrow slit in a large conducting plane, when a uniform plane is incident. The results for two polarizations are shown.

INTRODUCTION

IN attempts to measure, at microwave frequencies, the diffraction from a long slit in a large conducting plane, the most troublesome difficulties concern (a) generation of a suitable uniform plane wave, (b) uniform illumination in phase and amplitude of the diffraction screen, (c) elimination of reflections in the shadow zone, (d) suitable power levels in the shadow zone to permit detection, (e) sampling of the shadow zone fields without undue perturbation.

In the measurements here described all of the above difficulties, save complete illumination of screen and slit, were overcome except that to satisfy (d) the opening could not be reduced to less than 0.5 cm when the electric vector was parallel to the slit or to 0.75 cm for the electric vector perpendicular to the slit. The free-space wavelength was 3.18 cm.

MEASURING APPARATUS

Rather than conduct any experimentation out-of-doors, it was felt that suitable free-space conditions could be simulated in the microwave counterpart of the acoustical anechoic chamber. At the Ipswich Field Station of the Air Force Cambridge Research Center, a room approximately 25 feet by 12 feet, with its walls and ceiling completely covered by a black absorbing material, was used to make the measurements.

A diffraction screen was built which essentially divided the room into two chambers, the light and the shadow zones. This screen was constructed of a four foot square metallic section of 0.050 inch aluminum centered in a 2-foot border of the absorbing material. The metallic portion consisted of two equal pieces with provision for adjusting the width of the vertical slit between them from zero to three centimeters. The shadow edges of the slit were ground down to a knife edge in order that the condition of zero thickness sheet be more nearly met at the aperture. The absorbing border was included to preclude the possibility of any energy passing into the shadow zone other than through the aperture.

In order to illuminate the aperture, a horn-driven binormal lens fed by a 2K39 klystron at 9540 mc was placed in front of the screen. Phase plots of the lens indicate that cylindric beam of equiphase-front energy was incident thereon. The lens played a twofold role: (a) concentrating the energy, and (b) generating a constant phase-front wave.

The diffracted fields were sampled by a small electric dipole less than a half-wavelength long. The dipole was placed into the fields with its support rod oriented in the direction of propagation of the diffracted energy so that its effect upon either the electric or magnetic components was minimized. The probe was matched at its junction to the coaxial connecting cable with a double-stub coaxial tuner.

Measurements in space were accomplished by supporting the probe assembly on a dielectric rod which moved on a calibrated carriage bed by means of a hand-driven lead screw. Thus, radial variations were observed by merely setting the carriage and noting the indicator reading. Azimuthal variations were observed by mounting the entire carriage assembly on a properly centered rotating plate.

Indication of relative field amplitudes was observed

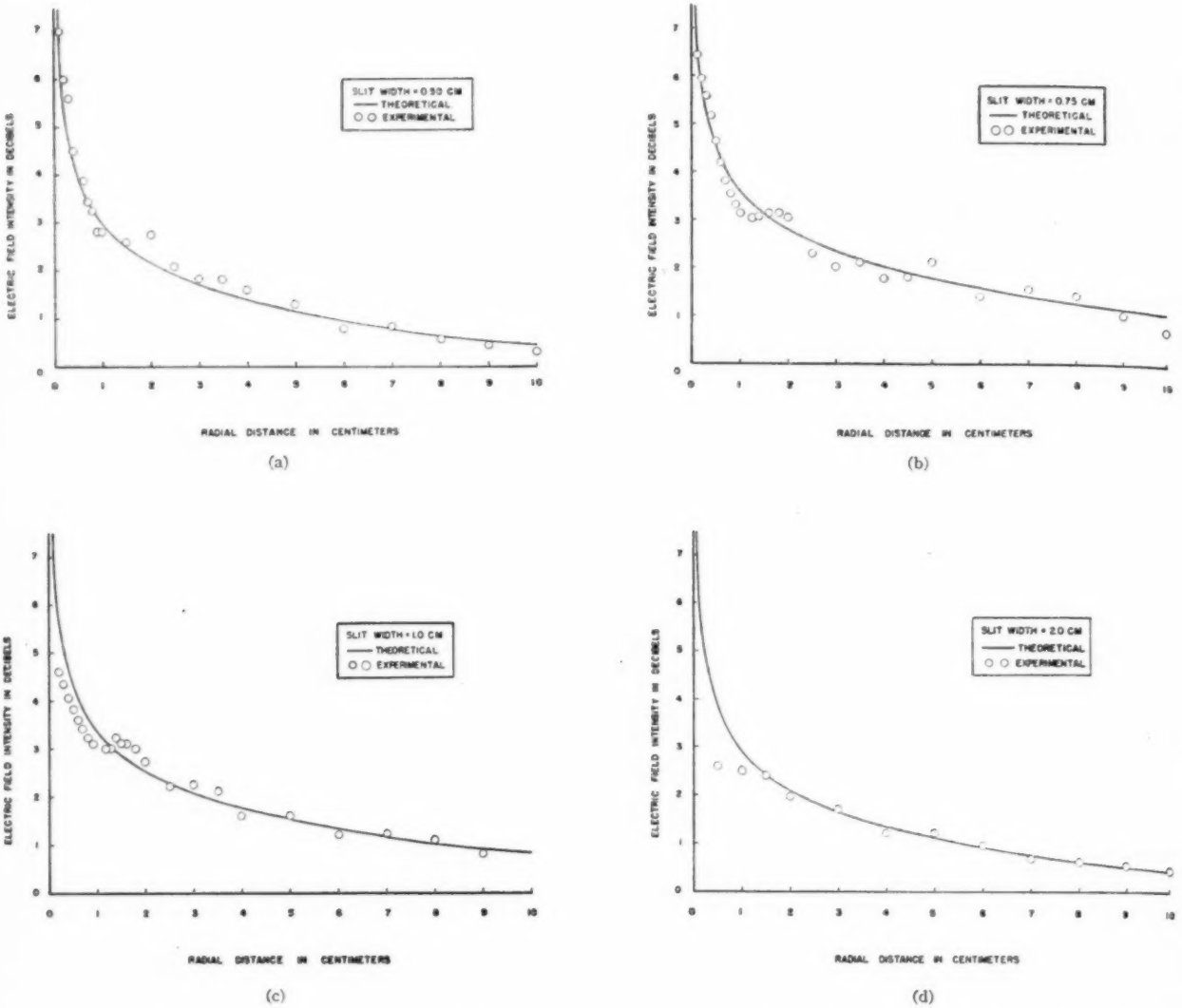


FIG. 1. Radial variations in electric field intensity taken normal to the diffraction screen for a vertically polarized incident wave.

on a Vectron model SA-10 spectrum analyzer. This instrument was chosen because of its extreme sensitivity as a receiver and because of its large (5 inch) oscilloscope face. Measurements were made by returning the appropriate spectral line on the analyzer back to its original amplitude at each sampled point in the field with a calibrated wave-guide-beyond-cutoff attenuator in the rf head of the analyzer. Because of inherent spectrum analyzer scintillations, it was not possible to read amplitudes to any greater accuracy than 0.25 db. For this reason, many of the plotted points represent an average of three independent readings.

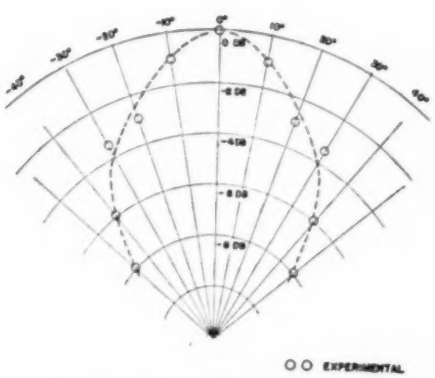


FIG. 2. Azimuthal variation in electric field intensity for a 1-cm slit with vertically polarized incident wave.

RESULTS OF MEASUREMENTS

The method of measurement made it possible to record only relative, rather than absolute values of field amplitudes. This introduces an arbitrary multiplying factor in the experimental variations obtained.

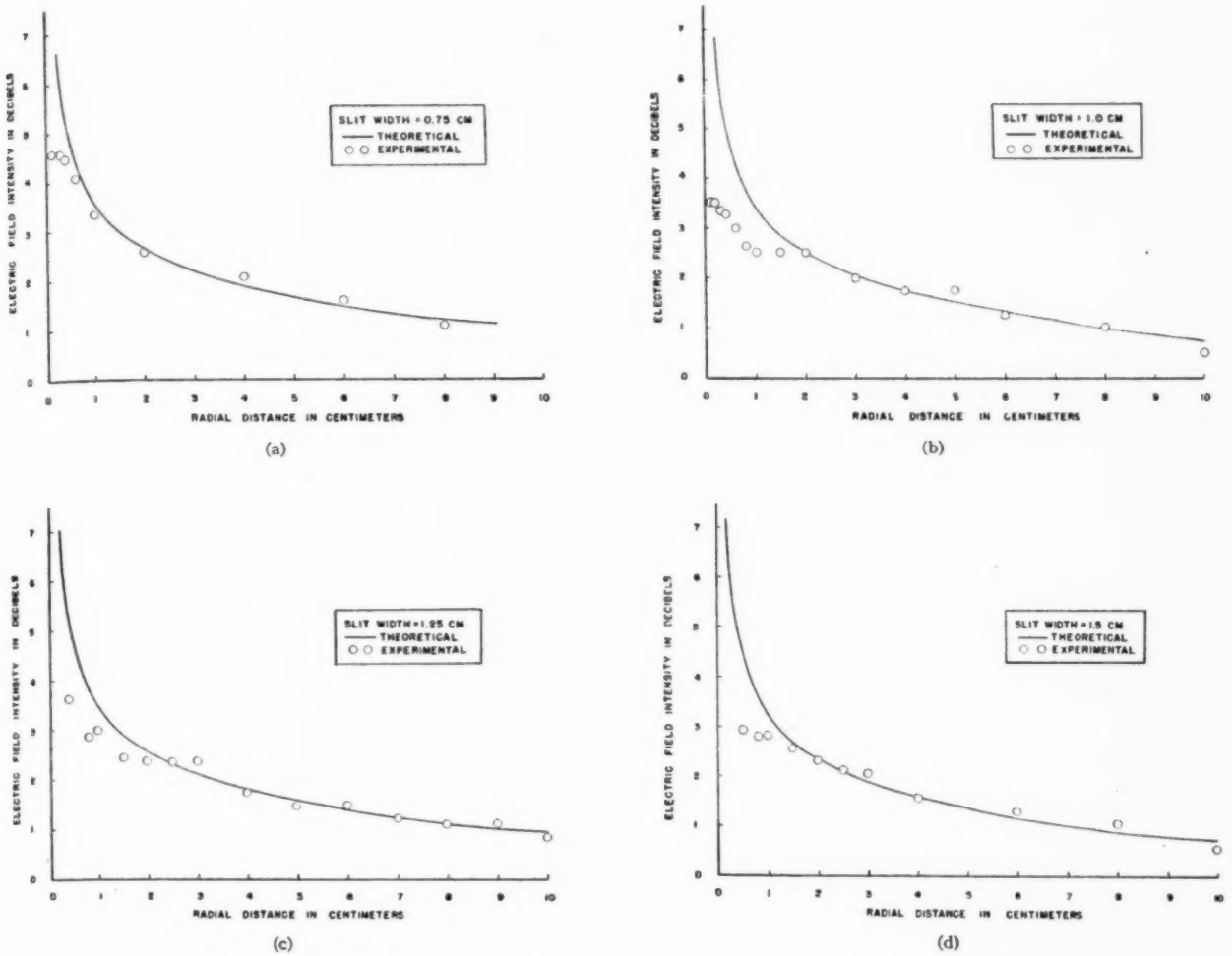


FIG. 3. Radial variations in electric field intensity taken normal to the diffraction screen for a horizontally polarized incident wave.

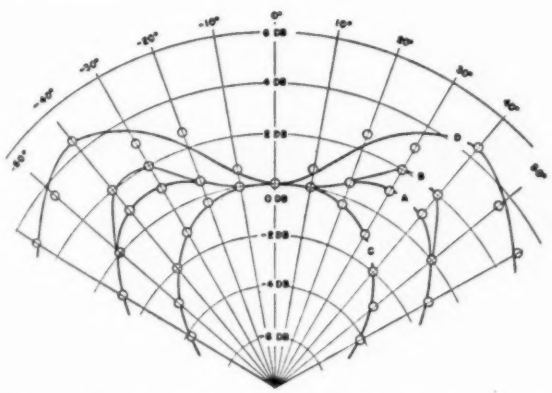


FIG. 4. Azimuthal variations in electric field intensity for a horizontal polarized incident wave. (Note: All maxima are normalized to 0 decibels.) (A) Slit width 1.5 cm. (B) Slit width 1.25 cm. (C) Slit width 1.0 cm. (D) Slit width 0.75 cm.

Results of the measurements are displayed in Figs. 1–4. The solid line in Figs. 1 and 3 is a plot of the magnitude of $H_1^{(2)}(\beta\rho)$.

Radial variations of Figs. 2 and 4 were taken at $\rho = 10$ cm at which radius the data were typical of that taken at other radii out to 25 cm. Curves for four slit widths are shown in Fig. 4 to illustrate more clearly the curious interference effect producing in three instances an unexpected maximum at $\theta = 35°$.

ACKNOWLEDGMENT

The above measurements were made possible by the cooperation of Dr. R. C. Spencer and his associates at the Antenna Laboratory of the Electronics Research Directorate of the Air Force Cambridge Research Center.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 2 FEBRUARY, 1955

Backscattering from Wide-Angle and Narrow-Angle Cones

LEOPOLD B. FELSEN
Microwave Research Institute, Polytechnic Institute of Brooklyn, Brooklyn 1, New York
(Received March 3, 1954)

Solutions are obtained for the diffraction of the waves radiated by scalar and vector point sources on the axis of a semi-infinite cone. The scalar problems are solved by the method of characteristic Green's functions to yield directly various alternative representations whose different convergence properties are discussed; the vector problem is solved by an application of spherical transmission line theory. To evaluate the plane wave scattering observed far from the cone tip, a highly convergent contour integral representation is selected and evaluated approximately for the special case of backscattering from cones having large and small apex angles. The results for the large-angle cone exhibit the transition from a backscattered spherical wave to a plane wave as the cone degenerates into an infinite plane.

1. INTRODUCTION

THE problems of the scattering of acoustic and electromagnetic waves by obstacles of simple geometry and composition have been among the first to be investigated in mathematical physics. Such simple scatterers as the sphere, the cylinder, the wedge, and the cone made their first appearance in the literature many years ago. Although mathematically rigorous solutions for these simple problems have been obtained for the case of simple (separable) boundary conditions, the subject is by no means exhausted because of the limited practical range of many of the formal results. The solutions are not given in closed form but rather in terms of infinite series or integrals so that the direct practical applicability of the formal results depends on the rapidity of convergence of a particular representation. The convergence properties of a given solution usually are favorable only for a limited range of values of the obstacle parameters, wavelength, and location of source and observation points. It is desirable, therefore, to have at one's disposal alternative forms of the result which exhibit different convergence properties.

A striking example of the utility of alternative representations presents itself in connection with the well-known problem of diffraction of the radiation from a vertical dipole source by a sphere. The classical solution in terms of modes (eigenfunctions) in the angular direction converges rapidly if the sphere radius is much smaller than the wavelength, but is very slowly convergent for large radius. By means of function-theoretic techniques this representation may be transformed into one exhibiting much more desirable convergence properties,[1] especially for the practically important case where the antenna and the receiver are both located on the surface of the sphere (earth). The new representation is a solution in terms of the modes in the radial domain.[2] A more direct and systematic approach for obtaining alternative representations of the solutions of this and other separable scalar diffraction problems is afforded by the method of characteristic Green's functions which has been employed by Marcuvitz in connection with scalar two-dimensional problems in spherical regions.[3] The characteristic Green's functions are analogous to voltages or currents on a transmission line; the alternative representations may be interpreted as arising from a transmission line formulation in different coordinates, i.e., transmission is assumed to take place along one or the other coordinate direction.

The characteristic Green's function approach is employed in this paper to obtain alternative representations for the two-dimensional scalar Green's functions for a cone with simple (Dirichlet or Neumann type) boundary conditions. The convergence properties of the various representations are described. A series solution in terms of angular modes (transmission in radial direction) is rapidly convergent when either the source point or the observation point is located near the cone tip, a radial mode (angular transmission) solution is useful for the evaluation of the far scattering from the tip of a narrow angle cone, and a mixed formulation is appropriate for the determination of the geometric-optical result of rays reflected from the sides of the cone. However, even the rapidly convergent representations present difficulties because of the complicated nature of the functions involved. A detailed discussion of various alternative representations for three-dimensional Green's functions and of their convergence properties in regions described simply in terms of spherical coordinates has been presented in a technical report.[4] Among the special examples given there, the radial mode solution alluded to above is included; its evaluation for plane wave scattering by a narrow-angle cone for off-axis or on-axis incidence and observation has also been carried out.[5] The present paper restricts itself to a consideration of the angular mode representation of the two-dimensional Green's

[1] G. N. Watson, Proc. Roy. Soc. (London) **95**, 63 (1918).
[2] A. Sommerfeld, *Partial Differential Equations in Physics* (Academic Press, Inc., New York, 1949), Ch. V, Appendix II, and Ch. VI, Appendix.
[3] N. Marcuvitz, Comm. in Pure and Appl. Math. 4, (1951).
[4] L. B. Felsen, "Alternative Field Representations in Regions Bounded by Spheres, Cones, and Planes," Microwave Research Institute, Polytechnic Institute of Brooklyn, Report R-359-54, PIB-293, January, 1954.
[5] L. B. Felsen, "Plane Wave Scattering by Small Angle Cones," Microwave Research Institute, Polytechnic Institute of Brooklyn, Report R-362-54, PIB-296, January, 1954.

function for the purpose of deriving approximate closed form results for the backscattering from cones with wide and narrow apex angles.

As implied above the angular mode series representation is slowly convergent for far zone plane wave scattering since both the source point and the observation point are located far from the cone tip. Thus, direct summation of the series is impractical. A more convenient representation is obtained by casting the solution into the form of a contour integral which exhibits the dependence on the large parameter kr ($k=2\pi/$wavelength) in such a form as to allow a direct asymptotic evaluation. The integrand, still given in the form of an infinite series, may be summed into a closed form result for backscattering along the axis from wide-angle and narrow-angle cones upon employing suitable approximations for the functions involved. For the wide-angle cone (apex angle $\varphi\approx\pi$) the integrand contains a double pole near a saddle point. The asymptotic evaluation of such an integral does not appear to have been discussed in the literature and is carried out in an Appendix. The result yields a backscattered wave which is spherical when φ is different from π, passes through a transition region as $\varphi\rightarrow\pi$, and becomes a plane wave when $\varphi=\pi$ (infinite plane). For the narrow-angle cone (φ small) the asymptotic evaluation is straightforward and leads to a backscattered wave which vanishes when $\varphi=0$ (free space). The solutions are carried out in Section II for the cases where the Green's function or its derivative vanishes on the cone surface (Dirichlet and Neumann type, respectively). The Neumann type Green's function is proportional to the acoustic pressure in the presence of a rigid obstacle.

In Section III the electromagnetic (vector) problem of the scattering by a perfectly conducting cone of the radiation from a transverse dipole source on the cone axis is solved by spherical transmission line theory. From this viewpoint, the cone surface is considered to constitute the wall of a spherical (conical) wave guide in which transmission takes place along the radial direction. The spherical transmission line approach has previously been discussed in detail[6–10] and may be applied directly to problems involving arbitrary source distributions.[8–10] A discussion of the vector Green's functions in various alternative representations is given in the above-mentioned report.[4] The restriction in this paper to the special case of a transverse dipole on the cone axis is made because the aim is the derivation of approximate closed form expressions for the plane wave backscattering along the cone axis. The spherical

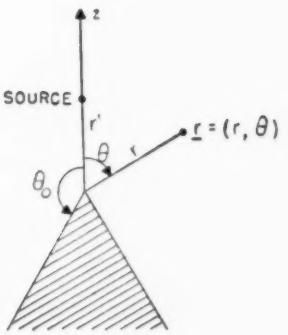


FIG. 1. Physical configuration.

transmission line representation has convergence properties similar to those of the angular mode solutions fo- the scalar problems. Approximate closed form results for the backscattering from wide-angle and narrowr angle cones are derived, and are found to agree, to first order, with the physical optics approximation of Spencer,[11] whose approximate validity in the intervening range has been verified by the experimental results of Sletten.[12]

The contents of this paper have been presented in a talk at the Symposium on Microwave Optics held at McGill University in Montreal, Canada, from June 22–25, 1953.[13] As a result of this conference it was found that a research group under K. M. Siegel at the Willow Run Research Center of the University of Michigan has also worked independently on the same problem.[14] Siegel *et al.* employ the approach of Hansen and Schiff[15] in deriving expressions for the plane wave backscattering in terms of "divergent" series which are then summed approximately by means of special summation techniques for wide-angle and narrow-angle cones. The pertinent technical reports[16,17] of the Willow Run research group have subsequently been made available to the author. It was found that the results for the spherical wave backscattering obtained here agree with those of Siegel *et al.*, to first order. An investigation of the next higher-order terms, contemplated by the author, has been dispensed with since such results have been obtained by the Willow Run research group. On the other hand,

[6] S. A. Schelkunoff, *Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York, 1943), Chap. 10 and 11.

[7] N. Marcuvitz, *Waveguide Handbook* (McGraw-Hill Book Company, Inc., New York, 1951), Secs. 1.8 and 2.8.

[8] Reference 3, Sec. 2b.

[9] L. B. Felsen, "Spherical Transmission Line Theory," Report R-253-51, PIB-194, January 1952, Microwave Research Institute, Polytechnic Institute of Brooklyn, Brooklyn, New York.

[10] L. B. Felsen and N. Marcuvitz, "Slot Coupling of Rectangular and Spherical Waveguides," J. Appl. Phys. **24**, (1953), Appendix.

[11] R. C. Spencer, "Back-Scattering from Conducting Surfaces" (AFCRC Report, April, 1951, Eq. (31)).

[12] C. J. Sletten, "Electromagnetic Scattering from Wedges and Cones," AFCRC Report, July, 1952, Fig. 13.

[13] L. B. Felsen, "Back-Scattering from Wide-Angle and Narrow-Angle Cones," Paper No. 5 in the symposium prepublications.

[14] K. M. Siegel, "Far Zone Assumption Put to Use," Symposium prepublications, Paper No. 8; C. E. Schensted, "Application of Summation Techniques to the Radar Cross-Section of a Cone," Symposium prepublication, Paper No. 9.

[15] W. W. Hansen and L. I. Schiff, "Theoretical Study of Electromagnetic Waves Scattered from Shaped Metal Surfaces," Quarterly Progress Report No. 4, Stanford University Microwave Laboratory, ATI 46568, September, 1948.

[16] K. M. Siegel and H. A. Alperin, "Studies in Radar Cross-Sections-III, Scattering by a Cone," Willow Run Research Center, University of Michigan Project MIRO, Report UMM-87, January, 1952.

[17] K. M. Siegel, *et al.*, "Studies in Radar Cross-Sections-IV, Comparison Between Theory and Experiment of the Cross-Section of a Cone," Project MIRO, Report UMM-92, February, 1953.

Siegel *et al.* do not have expressions valid in the transition region $\varphi \to \pi$ which have been determined here. A comparison between the methods employed by Siegel *et al.* and the author is given in Section II. The approach presented in this paper has the advantages of avoiding the manipulation of divergent series and of yielding quite naturally the scattering in the transition region as $\varphi \to \pi$.

2. SCALAR DIFFRACTION PROBLEMS

A. $\partial \mathcal{G}'/\partial\theta = 0$ on Cone Surface

(1) *Formulation of the Result*

The physical configuration of a source located at the point $\mathbf{r}'$ on the axis of symmetry of a semi-infinite cone is shown in Fig. 1. The cone apex is situated at the origin. The Green's function $\mathcal{G}'$ satisfies the inhomogeneous wave equation

$$(\nabla^2 + k^2)\mathcal{G}'(\mathbf{r},\mathbf{r}') = -\delta(\mathbf{r}-\mathbf{r}'), \quad k \text{ real} \tag{2.1}$$

with the following boundary conditions:

$$\mathcal{G}' \text{ finite at } r=0; \quad r(\partial\mathcal{G}'/\partial r - ik\mathcal{G}') \to 0 \text{ as } r \to \infty \tag{2.1a}$$

$$\partial\mathcal{G}'/\partial\theta = 0 \text{ at } \theta = \theta_0. \tag{2.1b}$$

The delta function is defined in the usual manner by

$$\delta(\mathbf{r}-\mathbf{r}') = 0, \quad \mathbf{r} \neq \mathbf{r}'; \quad \int_V dV' \delta(\mathbf{r}-\mathbf{r}') = 1, \quad \mathbf{r} \text{ in } V. \tag{2.2}$$

The Green's function $\mathcal{G}'$ may represent, to within a constant factor, the acoustic pressure for a point source outside a perfectly rigid cone, or the electromagnetic Hertz potential of a magnetic dipole directed along the axis of a perfectly conducting cone. Equation (2.1) may be written in the form

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + k^2\right) r\mathcal{G}'(\mathbf{r},\mathbf{r}') = -\frac{\delta(r-r')\delta(\theta-\theta')}{2\pi r' \sin\theta'}. \tag{2.3}$$

From a characteristic Green's function point of view, the two one-dimensional operator problems characteristic of the (separable) two-dimensional operator problem in the given $r\theta$ domain are analyzed separately to yield the characteristic Green's functions and the associated complete set of eigenfunctions. The solutions of the one-dimensional characteristic problems are then combined to yield directly alternative representations for the two-dimensional Green's function. For the problem under consideration the two one-dimensional characteristic problems in the r and θ domains are included among those discussed by Marcuvitz[18] and are summarized below [for a time dependence $\exp(-i\omega t)$].

Radial domain.—

$$\left(\frac{d^2}{dr^2} + k^2 - \frac{\lambda_1}{r^2}\right) G_r(r,r';\lambda_1) = -\delta(r-r'), \quad 0 < r < \infty, \tag{2.4a}$$

$$\mathrm{Re}(\lambda_1 + \tfrac{1}{4})^{\frac{1}{2}} > 0. \tag{2.4b}$$

Boundary conditions:

$$G_r \text{ finite at } r=0; \quad (dG_r/dr - ikG_r) \to 0, \quad r \to \infty. \tag{2.4c}$$

Solution for G_r:

$$G_r(r,r';\lambda_1) = \frac{i}{k}\mathrm{j}_\nu(kr_<)\mathrm{h}_\nu^{(1)}(kr_>), \quad \lambda_1 = \nu(\nu+1). \tag{2.5}$$

Completeness relation:

$$r'^2\delta(r-r') = \frac{1}{2\pi i}\oint_{C_1} G_r(r,r';\lambda_1)\,d\lambda_1 \tag{2.6a}$$

$$= \frac{1}{2\pi k}\int_{\epsilon-i\infty}^{\epsilon+i\infty} d\nu(2\nu+1)\mathrm{j}_\nu(kr)\mathrm{h}_\nu^{(1)}(kr'), \tag{2.6b}$$

where $\epsilon = -\frac{1}{2}+0$. In Eq. (2.6b) r and r' may be interchanged. The path C_1 in the complex λ_1 plane encircles the branch cut along the negative real λ_1 axis in the positive sense. The spherical Bessel functions employed in Eq. (2.5) are related to the ordinary Bessel functions as follows:

$$z_\nu(x) = (\pi x/2)^{\frac{1}{2}} Z_{\nu+\frac{1}{2}}(x), \tag{2.7}$$

where Z denotes any of the solutions of the Bessel equation. $r_<$ stands for r when $r<r'$, and for r' when $r>r'$; the converse holds for $r_>$. The transform theorem contained in Eq. (2.6) was first investigated by Kontorovitch and Lebedev.[19]

Angular domain.—

$$\left(\frac{d}{d\theta}\sin\theta\frac{d}{d\theta} + \lambda_2\sin\theta\right) G_\theta'(\theta,\theta';\lambda_2) = -\delta(\theta-\theta'), \quad 0 \leqslant \theta \leqslant \theta_0, \tag{2.8a}$$

$$\mathrm{Im}\,\lambda_2 \neq 0. \tag{2.8b}$$

Boundary conditions:

$$G_\theta' \text{ finite at } \theta=0; \quad dG_\theta'/d\theta = 0 \text{ at } \theta=\theta_0. \tag{2.8c}$$

Solution for G_θ':

$$G_\theta'(\theta,\theta';\lambda_2) = -\frac{\pi}{2}\frac{P_\mu(\cos\theta_<)[P_\mu(\cos\theta_>)(d/d\theta_0)P_\mu(-\cos\theta_0) - P_\mu(-\cos\theta_>)(d/d\theta_0)P_\mu(\cos\theta_0)]}{\sin\mu\pi\,(d/d\theta_0)P_\mu(\cos\theta_0)}, \tag{2.9}$$

[18] Reference 3, Example $4b_4$ (with $\lambda \to \lambda_1$) and $4a_2$ (with $\theta \to \pi-\theta$, $\theta_0 \to \pi-\theta_1$).
[19] M. J. Kontorovitch and N. H. Lebedev, J. Phys. (U.S.S.R.) **1**, 229–241 (1939).

where $\lambda_2=\mu(\mu+1)$.

Completeness relation:

$$\frac{\delta(\theta-\theta')}{\sin\theta'}=-\frac{1}{2\pi i}\oint_{C_2} G_\theta'(\theta,\theta';\lambda_2)\,d\lambda_2 \tag{2.10a}$$

$$=-\frac{1}{2}\csc^2\frac{\theta_0}{2}-\frac{\pi}{2}\sum_p \frac{(2p+1)P_p(\cos\theta)P_p(\cos\theta')(d/d\theta_0)P_p(-\cos\theta_0)}{\sin p\pi[(\partial^2/\partial\mu\partial\theta_0)P_\mu(\cos\theta_0)]_{\mu=p}}, \tag{2.10b}$$

$$(d/d\theta_0)P_p(\cos\theta_0)=0=P_p^1(\cos\theta_0), \quad p>0. \tag{2.10c}$$

The path C_2 encircles the positive real λ_2 axis in the positive sense. Since G_θ' is an even function of $(\mu+\frac{1}{2})$, its only singularities are poles located at $\mu=p$. The summation in Eq. (2.10b) is carried out over the positive zeros $\mu=p$ of $(d/d\theta_0)P_\mu(\cos\theta_0)$, where $P_\mu(\cos\theta)$ is the Legendre function. The constant term in the spectral representation (2.10b) has not been correctly taken into account in footnote reference 18.

With the knowledge of the solutions of the two one-dimensional characteristic problems, the solution of Eq. (2.3) may be written down immediately in various alternative forms.[20] Two such directly obtained alternative representations for the case where the source point is located on the cone axis ($\theta'=0$) are the following:

$$\mathcal{G}'(\mathbf{r};\mathbf{r}',0)=\frac{i}{4krr'}\left\{\frac{1}{\pi}\csc^2\frac{\theta_0}{2}j_0(kr_<)h_0^{(1)}(kr_>)-\sum_p \frac{(2p+1)P_p(\cos\theta)(d/d\theta_0)P_p(-\cos\theta_0)j_p(kr_<)h_p^{(1)}(kr_>)}{\sin p\pi[(\partial^2/\partial\mu\partial\theta_0)P_\mu(\cos\theta_0)]_{\mu=p}}\right\}, \tag{2.11a}$$

$$=-\frac{1}{8\pi krr'}\int_C d\mu \frac{(2\mu+1)j_\mu(kr_<)h_\mu^{(1)}(kr_>)}{\sin\mu\pi(d/d\theta_0)P_\mu(\cos\theta_0)}[P_\mu(\cos\theta)(d/d\theta_0)P_\mu(-\cos\theta_0) - P_\mu(-\cos\theta)(d/d\theta_0)P_\mu(\cos\theta_0)], \tag{2.11b}$$

where p has been defined in Eq. (2.10c). The path C encloses the positive real μ axis in the positive sense and may be deformed directly throughout that portion of the μ plane for which $\mathrm{Re}\,\mu>(-\frac{1}{2})$, since the integrand vanishes sufficiently rapidly at $|\mu|\to\infty$. The formal solutions presented in Eqs. (2.11) were first obtained by Carslaw[21] by a different, somewhat less direct approach. Equation (2.11a) may be shown to reduce correctly to the expected results for a point source in free space when $\theta_0=\pi$, and for a point source with its image (to account for the infinite plane) when $\theta_0=\pi/2$.

The series representation in Eq. (2.11a) is rapidly convergent if either the source point or the observation point is located near the cone tip, owing to the behavior of $j_\mu(x)\cong[x/(\mu+\frac{1}{2})]^{\mu+1}$ when $x\ll\mu$, $\mu\gg1$. When both the source point and the observation point are situated very far from the tip, the contour integral representation (2.11b) is appropriate to the derivation of the ray-optical result to yield rays reflected from the (curved) sides of the cone according to the laws of geometrical optics. The contour integral form is convenient, moreover, for the direct formulation of the scattering due to the cone, since the contribution of the second term inside the square brackets in Eq. (2.11b) may be shown to be exactly that of a point source in free space. Thus, the remaining expression represents the perturbation introduced by the cone. When the path C is deformed into the contour $\mu=(-\frac{1}{2})+ix$, $-\infty<x<\infty$, the radial mode solution obtains and lends itself readily to the evaluation of the plane wave scattering by the tip of a cone with small apex angle.[5] This paper deals with the summation of the series in Eq. (2.11a) into an approximate closed form result for the special, but important, case of plane wave backscattering from wide-angle and narrow-angle cones.

In order to determine explicitly the diffraction of the incident wave by the obstacle, it is necessary to present the final solution in a form which exhibits separately the contributions from the incident and scattered waves. If the backscattered wave is assumed to be spherical, the solution must take the form

$$\mathcal{G}'(r,0;\infty,0)\sim e^{-ikr}+f(k,\theta_0)e^{ikr}/r, \quad kr\gg1. \tag{2.12}$$

It is evident that $\mathcal{G}'(r,0;\infty,0)$ as represented in Eq. (2.11a) does not exhibit the separation demanded in Eq. (2.12). (However, as pointed out above, this separation is achieved by the angular transmission formulation.) To force this separation in Eq. (2.11a), the incident wave may be expanded in terms of the characteristic functions for the cone, and from the difference $[\mathcal{G}'-\exp(-ikr)]$, for $kr\gg1$, the expression for $f(k,\theta_0)$ is deduced. This approach was employed by Hansen and Schiff,[15] and followed by Siegel *et al.*,[16,17] in their investigation of the cone problem. In obtaining the asymptotic forms ($kr\to\infty$) for the pertinent series expansions these authors replaced, in effect, the spherical Bessel functions by their large argument approximation in every term of the series. A consequence of this approximation is a resulting representation for $f(k,\theta_0)$ in terms of a "divergent" series.

[20] Reference 3, Eqs. (5.1)–(5.4).
[21] H. S. Carslaw, Math. Ann., **75**, 133–147 (1914).

In an alternative approach, employed in this paper, the breakup into incident and scattered waves is not attempted in the series representation (2.11a). Instead, a different representation is obtained by replacing the spherical Bessel functions in Eq. (2.11a) by their Sommerfeld integral representation, which is valid for all values of order and argument, and by interchanging the order of summation and integration.[22] The resulting series is still uniformly convergent and may be approximated by suitable closed form expressions. For large values of kr, the remaining contour integral is evaluated by the method of saddle points, and the result exhibits directly the desired separation into incident and scattered waves. The scattered wave is determined by the nature of the integrand and is not restricted to be spherical as in Eq. (2.12). In particular, one traces out quite readily the transition region which the backscattered wave must undergo when the external cone angle $\theta_0 \to \pi/2$, i.e., the cone degenerates into an infinite plane. In the domain of existence of a backscattered spherical wave one obtains a series represenation for $f(k,\theta_0)$ which is uniformly convergent in view of an exponential damping factor, and whose closed form result is continued analytically into a range where the damping factor vanishes. In the approach of Hansen and Schiff, the damping factor is absent from the original series expression; as a consequence, the series is divergent.

The desired representation is obtained by substituting the asymptotic expression

$$h_p^{(1)}(kr') \sim \exp[kr' - (p+1)\pi/2], \quad kr' \to \infty, \tag{2.13}$$

and the Sommerfeld integral formula

$$j_p(kr) = \frac{1}{2\pi}(\pi kr/2)^{\frac{1}{2}} \int_W d\psi e^{ikr\cos\psi + i(p+1/2)(\psi - \pi/2)} \tag{2.14}$$

into Eq. (2.11a), and interchanging the orders of summation and integration. The interchange is permissible since $\operatorname{Im}\psi > 0$ for the path W chosen in Fig. 2, so that $\exp(ip\psi)$ provides exponential damping, and the series is uniformly convergent. Inside the shaded portions of the ψ plane, $\exp(ikr\cos\psi)$ vanishes exponentially at infinity. The resulting representation along the $\theta=0$ axis is the following:

$$\mathcal{G}'(r,0;\infty,0) \equiv \mathcal{G}'(r) = e^{i\pi/4}(\pi/2kr)^{\frac{1}{2}} \int_W d\psi e^{ikr\cos\psi} S'(\psi), \tag{2.15}$$

where

$$S'(\psi) = \frac{1}{2\pi}\csc^2\frac{\theta_0}{2} e^{i(\psi-\pi)/2} - \sum_p \left\{ \frac{(d/d\theta_0)P_p(-\cos\theta_0)}{\sin p\pi [(\partial^2/\partial\mu\partial\theta_0)P_\mu(\cos\theta_0)]_{\mu=p}} \right\} (p+\tfrac{1}{2}) \times e^{i(p+\frac{1}{2})(\psi-\pi)}. \tag{2.16}$$

In Eqs. (2.15) and (2.16) the factor $[\exp(ikr')]/4\pi r'$ has been set equal to unity to make $\mathcal{G}'$ the response to a unit intensity incident plane wave (see Appendix B). For large values of kr, the integral in Eq. (2.15) may be evaluated asymptotically provided that $S'(\psi)$ is known in closed form. Approximate closed form expressions for S' will now be derived for wide-angle and narrow-angle cones.

(2) *Approximations for a Wide-Angle Cone*: $\theta_0 \approx \pi/2$

In order to derive a closed form result for the series in Eq. (2.16) it is necessary first to obtain an analytic expression for the summation indices p. Such an expression, in general, is very complicated. The large values of p, however, may be found directly from the asymptotic form of the Legendre function[23]

$$P_\mu^\eta(\cos\theta) \sim \mu^\eta (2/\pi\mu\sin\theta)^{\frac{1}{2}} \times \cos\left[(\mu+\tfrac{1}{2})\theta - \frac{\pi}{4} + \frac{\eta\pi}{2}\right] + O(\mu^{\eta-3/2}), \tag{2.17}$$

which is valid for $|\mu| \gg |\eta|$, $\mu\sin\theta \gg 0$. Thus, the large values of p are given approximately by the zeros of $P_\mu^1(\cos\theta_0) = (d/d\theta_0)P_\mu(\cos\theta_0)$ as follows:

$$p \approx (m+\tfrac{1}{4})\pi/\theta_0 - \tfrac{1}{2}, \quad m = \text{large positive integer}. \tag{2.18}$$

It is readily found by employing Eqs. (2.17) and (2.18) that, to first order, the group of terms inside the braces in Eq. (2.16) is equal to $(-1/\theta_0)$. (It can also be checked that this approximation is good to second order, for large values of p.) Moreover, for the special case $\theta_0 = \pi/2$ the value $(-1/\theta_0) = -(2/\pi)$ is *exact* for *all* values of p, and all the zeros are given exactly by $p = 2m$, $m = 1, 2, 3\cdots$.[24] It is reasonable to expect, therefore, that for $\theta_0 \approx \pi/2$, the above expressions are approximately valid over the entire range. Under these

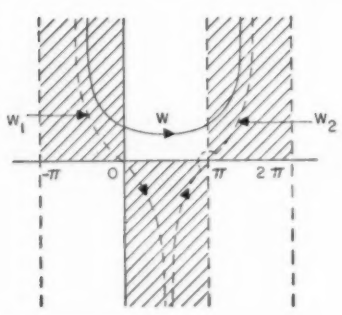


FIG. 2. Paths of integration in the ψ plane.

[22] This representation is analogous to that employed by MacDonald in the solution of the wedge problem [see H. M. MacDonald, *Electromagnetism* (G. Bell and Sons, London, 1934), p. 79].

[23] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge University Press, Cambridge, 1931), p. 303.

[24] See reference 23, p. 217, Eq. (47).

assumptions, $S'(\psi)$ is given approximately by

$$S'(\psi)\approx\frac{1}{\theta_0}\sum_{m=0}^{\infty}(p+\tfrac{1}{2})e^{i(p+\frac{1}{2})(\psi-\pi)}$$

$$=\frac{1}{\theta_0}\frac{d}{id\psi}\left[e^{i(\alpha/4)(\psi-\pi)}\sum_{m=0}^{\infty}(e^{i\alpha(\psi-\pi)})^m\right] \tag{2.19a}$$

$$=-\frac{\alpha}{4\theta_0}\frac{e^{-i\frac{1}{4}\alpha(\psi-\pi)}[\frac{1}{4}+\frac{3}{4}e^{i\alpha(\psi-\pi)}]}{\sin^2[(\alpha/2)(\psi-\pi)]},\quad \alpha=\pi/\theta_0, \tag{2.19b}$$

where the $p=0$ term in Eq. (2.16) has also been included in the sum. Since the series in Eq. (2.19a) are uniformly convergent ($\mathrm{Im}\,\psi>0$), the interchange of differentiation and summation is permissible; also the summation of the second series as a power series is valid. The closed form result in Eq. (2.19b) is valid for $\mathrm{Im}\,\psi<0$ as well so that the expression can be continued analytically into the lower half of the ψ plane.

(3) *Approximations for a Narrow-Angle Cone*: $\theta_0\approx\pi$

The zeros p for this case are given approximately by[25]

$$p\approx m+m(m+1)(\pi-\theta_0)^2/4=m+\Delta_m,\quad \theta_0\to\pi, \tag{2.20}$$

where $m=1, 2, 3\cdots$. Also,[26]

$$P_p(\cos\theta_0)\approx(2/\pi)\sin p\pi\ln[(\pi-\theta_0)/2],\qquad p(\pi-\theta_0)/2\ll 1, \tag{2.21}$$

and

$$P_p(-\cos\theta_0)\approx 1-p(p+1)\sin^2[(\pi-\theta_0)/2]. \tag{2.21a}$$

Upon substituting Eqs. (2.20) and (2.21) into Eq. (2.16) one obtains to $0[(\pi-\theta_0)^2]$

$$S'(\psi)\approx-\frac{1}{\pi}[\tfrac{1}{2}e^{i(\psi-\pi)/2}+S_1(\psi)]+\frac{1}{\pi}\left(\frac{\pi-\theta_0}{2}\right)^2[\tfrac{1}{2}e^{i(\psi-\pi)/2}+S_2(\psi)], \tag{2.22}$$

where

$$S_1(\psi)=\frac{d}{id\psi}\sum_{m=1}^{\infty}e^{i(m+\frac{1}{2})(\psi-\pi)}e^{i\Delta_m(\psi-\pi)}, \tag{2.23a}$$

$$S_2(\psi)=2\sum_{m=1}^{\infty}(m+\tfrac{1}{2})^2\,e^{i(m+\frac{1}{2})(\psi-\pi)}. \tag{2.23b}$$

As before, the interchange of summation and differentiation in Eq. (2.23a) is permissible. In order to express $S_1(\psi)$ in terms of powers of $(\pi-\theta_0)^2$, the second exponential in the series is represented by the first two terms of its power series expansion. This implies not only that $\Delta_m\ll 1$, but also that $\Delta_m(\psi-\pi)\ll 1$. If $\Delta_m\ll 1$, the latter condition is satisfied for the contributing range of the integral in Eq. (2.15), since its maximum contribution arises from the neighborhood of the saddle points at $\psi=0,\ \pi$. The condition, $p(\pi-\theta_0)/2\ll 1$, restricts the applicability of the approximate representation in Eq. (2.22) to the first terms in the series. However, for $p\gg kr$, the terms in the series representation in Eq. (2.11a) are negligible, owing to the behavior of $j_p(kr)$ in that range. Therefore, if the restriction $1/(\pi-\theta_0)\gg kr\gg 1$ is added, the approximation $\Delta_m\ll 1$ applies correctly to the contributing range of the series. The approximate closed form expression for S' is then found to be, to $0[(\pi-\theta_0)^2]$,

$$\pi S'(\psi)\approx\left\{\frac{d}{id\psi}-\left(\frac{\pi-\theta_0}{2}\right)^2\left[\left(3\frac{d^2}{d\psi^2}+\frac{1}{4}\right)+(\psi-\pi)\left(\frac{d^3}{d\psi^3}+\frac{1}{4}\frac{d}{d\psi}\right)\right]\right\}S_3(\psi), \tag{2.24}$$

with

$$S_3(\psi)=\sum_{m=0}^{\infty}e^{i(m+\frac{1}{2})(\psi-\pi)}=(i/2)\csc[(\psi-\pi)/2]. \tag{2.24a}$$

The closed form result in Eq. (2.24) may be continued analytically into the lower half of the ψ plane.

(4) *Asymptotic Evaluation of the Contour Integral*

With the knowledge of the approximate closed form results for $S'(\psi)$, the integral in Eq. (2.15) may be evaluated asymptotically for large values of kr. The integrand contains pertinent saddle points at $\psi=0,\ \pi$, and the path of integration W may be deformed into the steepest descent paths W_1 and W_2 through the saddle points (see Fig. 2.2). For the wide-angle cone, the function S' given in Eq. (2.19b) has double poles at

$$\psi_1=\pi-2\theta_0\quad\text{and}\quad\psi_2=\pi. \tag{2.25}$$

It is assumed in the following that $\theta_0\gtrsim\pi/2$ so that no singularities lie on the real ψ axis between 0 and π. Therefore, the deformation of W into W_1 and W_2 can be carried out directly. (The case $\theta_0<\pi/2$ may be treated in the same manner except that the residue at the pole $\psi_1>0$ must be taken into account.)

The pole ψ_2 coincides with the saddle point at $\psi=\pi$. The asymptotic evaluation of integrals whose integrands have multiple poles at a saddle point is known[27] so that the result for the integration over W_2 may be written down directly. If, for a given value of $kr\gg 1$, θ_0 is such that $(2kr)^{\frac{1}{2}}|\cos\theta_0|\gg 1$, the double pole ψ_1 may be considered to be located far from the saddle point at $\psi=0$, and the asymptotic evaluation of the integral over the contour W_1 is straightforward as well. However, for $\theta_0\to\pi/2$ (our approximations are most valid in that range), the pole moves so close to the saddle point that its influence must be taken into account explicitly. The asymptotic evaluation of an integral whose integrand has a double pole near a saddle point is carried out in

[25] See reference 23, p. 408.
[26] See reference 23, pp. 225, 188.
[27] See B. L. Van der Waerden, Appl. Sci. Research **B2**, 191–217 (1951).

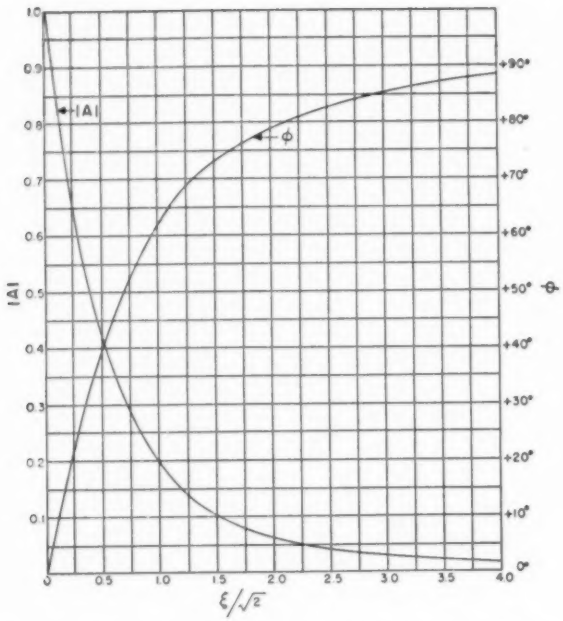


FIG. 3. Plot of $A=|A|e^{i\phi}=1-\xi\exp[-i(\xi^2+\pi/4)]\int_{-i\xi^2}^{\infty}dx\times e^{-x}x^{-\frac{1}{2}}$. ($A$ is the back-scattered amplitude for small values of $\xi=(2kr)^{\frac{1}{2}}|\cos\theta_0|$.)

Appendix A. The final first-order result is the following:

$$\mathcal{G}'(r)\sim e^{-ikr}+e^{ikr}\sin\theta_0\left\{1-\xi\exp\left[-\left(\frac{i\pi}{4}+i\xi^2\right)\right]\times\int_{-i\xi^2}^{\infty}dx\,e^{-x}x^{-\frac{1}{2}}+\cdots\right\},\quad \xi\text{ small},\qquad(2.26a)$$

$$\sim e^{-ikr}+\frac{i}{4(\theta_0-\pi/2)^2}\frac{e^{ikr}}{kr},\quad \xi\gg1,\quad \theta_0\approx\pi/2,\qquad(2.26b)$$

where $\xi=(2kr)^{\frac{1}{2}}|\cos\theta_0|$, $kr\gg1$. The integral appearing inside the braces in Eq. (2.26a) has been tabulated. A plot of the magnitude and phase of the terms inside the braces for different values of ξ is shown in Fig. 3.

The $\exp(-ikr)$ term in Eqs. (2.26) represents the incident plane wave along the z axis; it arises from the contribution of the saddle point at $\psi=\pi$. The remaining term represents the backscattered wave and is due to the contribution from the saddle point at $\psi=0$. When $\theta_0=\pi/2$, Eq. (2.26a) yields correctly the backscattered plane wave appropriate to reflection from an infinite plane. As θ_0 deviates from $\pi/2$, the backscattered wave passes through a transition region until the domain $\xi\gg1$ is reached in which the wave behaves like a spherical wave emanating from the cone tip. Detailed expressions necessary to trace the backscattered wave through all values of ξ are given in Appendix A. As ξ decreases toward zero a transition occurs because the z axis is located in a border region of the domain containing the rays reflected from the sides of the cone according to the laws of geometrical optics. These rays are confined to the region $\theta_1>2\theta_0-\pi$, i.e., $\theta_1\to0$ as $\theta_0\to\pi/2$.

For $\theta_0\approx\pi$, the function $S'(\psi)$ is given in Eqs. (2.24). Its explicit evaluation is straightforward, but somewhat lengthy, and is not reported here. It is found that there are no singularities near the saddle point at $\psi=0$ so that the integration over the path W_1 is straightforward. At $\psi=\pi$, the first term on the right-hand side of Eq. (2.24) has a double pole and may be evaluated directly as well. The remaining terms, multiplied by $(\pi-\theta_0)^2$, are of $0[1/(\psi-\pi)]$ at $\psi=\pi$; their contribution leads to higher order terms in kr and is neglected to first order. The evaluation of the integral over the path W_2 yields the incident wave only and serves as a check on the correctness and self-consistency of the first order result. It is found that, to $0[(\pi-\theta_0)^2]$, $\mathcal{G}'$ is given approximately by (see also reference 5)

$$\mathcal{G}'(r)\sim e^{-ikr}+\frac{i}{2}\left(\frac{\pi-\theta_0}{2}\right)^2\frac{e^{ikr}}{kr},\quad kr\gg1.\qquad(2.27)$$

The factor multiplying $[\exp(ikr)]/kr$ in Eqs. (2.26b) and (2.27) is equal to $\pi S'(0)$, where $S'(\psi)$ is given exactly in Eq. (2.16), and is also equal to $kf(k,\theta_0)$ in Eq. (2.12). It has therefore been proved that the backscattering for $\xi\gg1$ is given asymptotically by $\pi S'(0)/k$, provided that the saddle point at $\psi=\pi$ contributes only to the incident wave (this has been proved here only to first order, but must be valid in general since this saddle point gives rise to incoming waves). This same conclusion had been reached by Hansen and Schiff, and Siegel *et al.*,[15–17] via the procedure already alluded to [remarks following Eq. (2.12)]. The series repre-

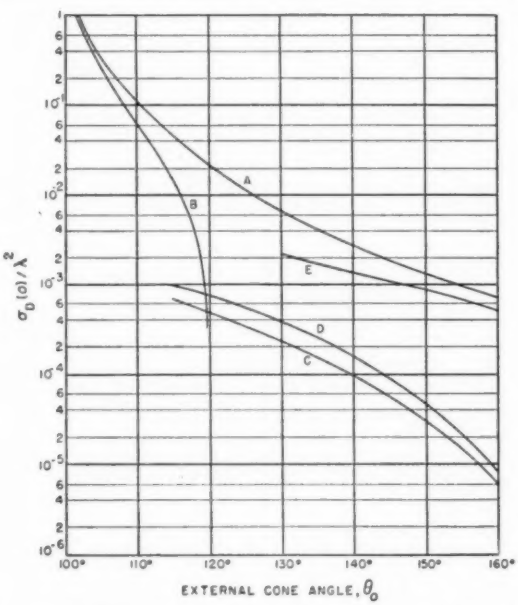


FIG. 4. Differential backscattering cross section (scalar problems). Neumann problem: A—1st order ($\theta_0\approx\pi/2$); B—2nd order ($\theta_0\approx\pi/2$) (Siegel *et al.*); C—1st order ($\theta_0\approx\pi$); D—2nd order ($\theta_0\approx\pi$) (Siegel *et al.*). Dirichlet problem: A—1st order ($\theta_0\approx\pi/2$); E—1st order ($\theta_0\approx\pi$).

sentation for $S'(\psi)$ in Eq. (2.16) is divergent when $\psi=0$ but uniformly convergent if $\operatorname{Im}\psi>0$. Thus, the direct summation procedure should be carried out on $S'(\psi)$ with $\operatorname{Im}\psi>0$, and the transition $\psi\to 0$ be performed on the closed form result.

The divergent series $S'(0)$ has been summed directly by Siegel *et al.*,[14–17] by means of special summation techniques for the cases of wide-angle and narrow-angle cones. (As mentioned in the Introduction, this work was carried out independently by Siegel *et al.*, and the author.) Since it can be verified that the summation techniques yield the same result as taking $\lim_{\psi\to 0}S'(\psi)$, the final expressions obtained by the two methods should be identical. The (second order) results for the differential backscattering cross section found by Siegel *et al.* are the following [see Eq. (2.12)][28]

$$\sigma_D(0)=|f(k,\theta_0)|^2=\frac{\lambda^2}{64\pi^2\cos^4\theta_0}(1-4\cos^2\theta_0+\cdots),\qquad \theta_0\approx\pi/2, \quad (2.28a)$$

$$=\frac{\lambda^2\gamma^2}{16\pi^2}\left[1+\left(4\ln\frac{1}{\gamma}-2\right)\gamma+\cdots\right],$$

$$\gamma=\sin^2\left(\frac{\pi-\theta_0}{2}\right),\qquad \theta_0\approx\pi, \quad (2.28b)$$

where $\lambda=2\pi/k$ is the applied wavelength. It is readily checked that the values of $\sigma_D(0)$ computed from Eqs. (2.26b) and (2.27) agree with the first-order terms in Eqs. (2.28a) and (2.28b), respectively. In the derivation of the expressions in Eqs. (2.28), Siegel *et al.* carried out the approximations of the Legendre functions and of the zeros p to the next higher order.

Although the solutions in Eqs. (2.26)–(2.28) are valid only when $\theta_0\approx\pi/2$ and $\theta_0\approx\pi$, it is instructive to plot their behavior over the intermediate range to demonstrate the degree of overlap between approximations applying at the end points. Such a plot is shown in Fig. 4 for the differential backscattering cross section. It appears that the first-order solutions of Eqs. (2.26) and (2.27) yield values in the intervening range which are too large and too small, respectively. The second-order solutions in Eqs. (2.28) have a common point near $\theta_0=120°$; but, since the expression for σ_D in Eq. (2.28a) is negative when $\theta_0>120°$, the significance of the cross-over point is questionable. It is difficult, therefore, to draw conclusions on the behavior of the correct result in the intervening range.

B. $\mathcal{G}=0$ on Cone Surface

The physical configuration for this problem is the same as that shown in Fig. 2.1. The Green's function $\mathcal{G}$ satisfies Eq. (2.1) with the boundary conditions in the radial domain given by Eq. (2.1a). However, in the θ domain,

$$\mathcal{G}(\mathbf{r},\mathbf{r}')=0 \text{ at } \theta=\theta_0. \quad (2.29)$$

The Green's function $\mathcal{G}$ so defined is proportional, for example, to the Hertz potential for an electric dipole located at $\mathbf{r}=(r',0)$ and directed along the cone axis.

In view of the unchanged boundary conditions in the radial domain, the characteristic problem for this case is identical with that in Part A of this section. The angular characteristic problem may be solved in analogy with the method employed in Part A. The result may be expressed as follows:

$$\mathcal{G}(\mathbf{r};\mathbf{r}',0)=-\frac{i}{4krr'}\sum_q(2q+1)\frac{P_q(\cos\theta)P_q(-\cos\theta_0)}{\sin q\pi[(\partial/\partial\nu)P_\nu(\cos\theta_0)]_{\nu=q}}\mathrm{j}_q(kr_<)\mathrm{h}_q^{(1)}(kr_>), \quad (2.30a)$$

$$=\frac{-1}{8\pi krr'}\int_C d\nu(2\nu+1)\frac{P_\nu(\cos\theta)P_\nu(-\cos\theta_0)-P_\nu(-\cos\theta)P_\nu(\cos\theta_0)}{\sin\nu\pi P_\nu(\cos\theta_0)}\mathrm{j}_\nu(kr_<)\mathrm{h}_\nu^{(1)}(kr_>). \quad (2.30b)$$

The summation indexes q are determined by

$$P_q(\cos\theta_0)=0. \quad (2.31)$$

The path C encloses the positive real ν axis in the positive sense. The convergence properties of Eqs. (2.30) are analogous to those of the corresponding representations in Eqs. (2.11). When $\theta_0=\pi$, Eq. (2.30a) yields the result for a point source in free space, while for $\theta_0=\pi/2$, one obtains a point source with its negative image.

For the plane wave backscattering observed far from the cone tip, the series solution in Eq. (2.30a) is converted into the form given in Eq. (2.15) with the following expression for $S(\psi)$ (instead of $S'(\psi)$):

$$S(\psi)=-\sum_q\frac{(q+\frac{1}{2})P_q(\cos\theta)P_q(-\cos\theta_0)e^{i(q+\frac{1}{2})(\psi-\pi)}}{\sin q\pi[(\partial/\partial\nu)P_\nu(\cos\theta_0)]_{\nu=q}},\qquad \operatorname{Im}\psi>0. \quad (2.32)$$

As before, the series may be summed approximately into a closed-form result for wide-angle and narrow-angle cones. The approximations for the wide-angle cone are obtained from Eq. (2.17), while those for the narrow-angle cone are derived by use of Eqs. (2.21) and the formula for the zeros

$$q\approx n+\delta,\quad 1/\delta=2\ln[2/(\pi-\theta_0)],\quad \theta_0\approx\pi,[24] \quad (2.33)$$

[28] Reference 17, Eq. (V-28).

where $n=0, 1, 2, \cdots$. It is found that

$$S(\psi)\approx[S'(\psi)]^*, \quad \theta_0\approx\pi/2 \tag{2.34a}$$

$$\approx -\frac{(1+2\delta)+(1-2\delta)e^{i(\psi-\pi)}}{8\pi\sin^2[(\psi-\pi)/2]}e^{i(2\delta-1)(\psi-\pi)/2}, \quad \theta_0\approx\pi. \tag{2.34b}$$

The expression for $S(\psi)$ in Eq. (2.34a) is the complex conjugate of that given for $S'(\psi)$ in Eq. (2.19b). The asymptotic evaluation of the integral in Eq. (2.15) for $kr\gg1$ proceeds as before. The final first-order result is given by

$$\mathcal{G}(r)\sim e^{-ikr}-e^{ikr}\sin\theta_0\{\ \}, \quad \xi \text{ small}, \tag{2.35a}$$

$$\sim e^{-ikr}-\frac{i}{4(\theta_0-\pi/2)^2}\frac{e^{ikr}}{kr}, \quad \xi\gg1, \quad \theta_0\approx\pi/2, \tag{2.35b}$$

$$\sim e^{-ikr}-\frac{i\delta}{2}\frac{e^{ikr}}{kr}, \quad kr\gg1, \tag{2.35c}$$

where $\xi=(2kr)^{\frac{1}{2}}|\cos\theta_0|$. The expression inside the braces in Eq. (2.35a) is identical with that in Eq. (2.26a).

The differential backscattering cross section computed from Eqs. (2.35b) and (2.35c) is plotted in Fig. 4. It is seen from Fig. 4 that the two curves approach one another in the region $\theta_0\approx150°$ so that it appears probable that curve A predicts the order of magnitude of the differential backscattering cross section correctly in the extended range $\theta_0<150°$.

3. VECTOR DIFFRACTION PROBLEM

A. Spherical Transmission Line Formulation

In Section II we have been concerned with the scalar problem of the diffraction of the waves radiated by a scalar source or an axial vector dipole source located on the axis of a semi-infinite cone. In this section attention is focused on the diffraction of the waves radiated by a transverse vector dipole source situated on the axis of a perfectly conducting cone. As distinct from the case of the axial dipole, the electromagnetic fields for the latter configuration may not be derived from a single scalar potential function satisfying the scalar wave equation. Instead, the vector character of the problem must be explicitly taken into account.

The vector diffraction problem under discussion is formulated quite readily in terms of spherical transmission line theory. From this viewpoint, the perfectly conducting cone surface is considered to constitute the wall of a spherical (conical) wave guide in which propagation takes place along the radial direction. The wave guide may be represented by an infinite number of spherical transmission lines each of which is excited, in general, by voltage and current generators whose strength depends on the sources in the field. The spherical transmission line problem need be solved only for a single typical transmission line, and the total solution is obtained by adding the contributions from all the lines. The transverse (θ,Φ) variation of the electromagnetic fields appears only in the form of transverse vector mode functions which are determined solely by the physical geometry of the wave guide and do not depend on the sources in the field. Once the transverse mode functions have been tabulated for a number of different geometries the solution of the electromagnetic field problem in such geometries reduces to that of a typical transmission line problem involving the pertinent source distribution. This section deals only with the special case of a transverse electric dipole located on the cone axis. It should be stressed, however, that the formal solution for an *arbitrary* source can be obtained just as readily from the spherical transmission line formalism as the special case discussed below.[4] The restriction to the special case is made here because the ultimate aim is the derivation of approximate closed form expressions for plane wave backscattering from wide-angle and narrow-angle cones, which can be carried out fairly simply only for the case of incidence along the cone axis.

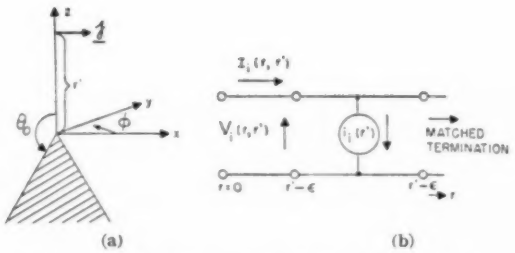


FIG. 5(a). Physical configuration. (b). Network representation for a typical spherical mode.

The properties of spherical transmission lines have been discussed in the literature.[6–10] It is found that the transverse electric current source in Fig. 5(a) may be represented by an equivalent current generator inserted into the ith spherical transmission line at the point $r=r'$ as shown in Fig. 5(b). The source excites both $E(TM)$ and $H(TE)$ modes in the spherical wave guide so that a double set of transmission lines is required. Each transmission line extends from the origin $r=0$ to infinity, giving rise to a standing-wave behavior for $r<r'$ and a traveling-wave behavior for $r>r'$. In particular, for a typical E-mode transmission line,[29]

$$\begin{aligned} I_i'(r,r') &= ji_i(r')\mathrm{j}_n(kr)\mathrm{h}_n'^{(2)}(kr'), \quad r<r' \\ &= ji_i(r')\mathrm{h}_n^{(2)}(kr)\mathrm{j}_n'(kr'), \quad r>r', \end{aligned} \tag{3.1a}$$

$$V_i'(r,r')=-\zeta i_i'(r')\mathrm{j}_n'(kr_<)\mathrm{h}_n'^{(2)}(kr_>), \quad \zeta=(\mu/\epsilon)^{\frac{1}{2}}. \tag{3.1b}$$

I_i' and V_i' are the E mode voltage and current, respectively, excited by the source i_i in the ith E-mode line. The subscript i stands for the double index (m,n) and ζ is the characteristic impedance of free space. In accordance with engineering practice, a time dependence

[29] Reference 10, Eqs. (A7) with $v_i'(r')=0$.

$\exp(+j\omega t)$ is assumed in the derivation of Eqs. (3.1) [$j=(-1)^{\frac{1}{2}}$]. The prime on the spherical Bessel functions denotes the derivative with respect to the argument. For the H-mode transmission lines one obtains similarly

$$\begin{aligned} I_i''(r,r') &= -ji_i''(r')\mathrm{j}_n'(kr)\mathrm{h}_n^{(2)}(kr'), \quad r<r' \\ &= -ji_i''(r')\mathrm{h}_n'^{(2)}(kr)\mathrm{j}_n(kr'), \quad r>r', \end{aligned} \tag{3.2a}$$

$$V_i''(r,r') = -\zeta i_i''(r')\mathrm{j}_n(kr_<)\mathrm{h}_n^{(2)}(kr_>). \tag{3.2b}$$

The double primes on V_i, I_i, and i_i denote H-mode quantities. In conformity with the network picture in Fig. 5(b) the mode currents are discontinuous, while the mode voltages are continuous, across the current source at $r=r'$.

The intensity of the current generator i_i is related to the intensity of the current source $\mathbf{J}$ as follows:[30]

$$i_i(r) = \int\int_{\Omega} r\mathbf{J}(r,\theta,\phi)\cdot\mathbf{e}_i(\theta,\phi)\sin\theta d\theta d\phi, \tag{3.3}$$

where $\mathbf{e}_i$ is the transverse electric vector mode function and the integration extends over the transverse domain Ω in which $\mathbf{J}$ is different from zero. Equation (3.3) is valid for E and H modes if $\mathbf{e}_i$ is replaced by $\mathbf{e}_i'$ and $\mathbf{e}_i''$, respectively. For the case under consideration, the current source $\mathbf{J}$ consists of an electric dipole which may be taken to be oriented parallel to the x axis in Fig. 3.1a without loss of generality. Thus,

$$\mathbf{J}(\mathbf{r}) = J\frac{\delta(r-r')\delta(\theta)\delta(\phi)}{r'^2\sin\theta'}\boldsymbol{\theta}_0, \tag{3.4}$$

so that

$$i_i(r) = (J/r')\boldsymbol{\theta}_0\cdot\mathbf{e}_i(0,0)\delta(r-r') = i_i(r')\delta(r-r'). \tag{3.5}$$

The orthonormal transverse vector mode functions $\mathbf{e}_i'$ and $\mathbf{e}_i''$ may be expressed in terms of scalar functions which have been determined previously.[31] In view of the special location and orientation chosen for the dipole source in Fig. 5(a), not all of the possible spherical modes are excited. It can be shown that the only modes excited are E modes with even symmetry in ϕ and H modes with odd symmetry in ϕ and that, moreover, the ϕ variation is restricted to $\sin\phi$ and $\cos\phi$. The latter property could also have been deduced from the fact that the unperturbed field (in free space) of a horizontal dipole located on the z axis and directed parallel to the x axis has such a ϕ variation.[32] Since the conical obstacle does not disturb the symmetry in ϕ, the total field must have this simple variation as well.

From the knowledge of the spherical mode voltages and currents and the vector mode functions, the formal solution in terms of spherical modes for the electromagnetic fields everywhere may be written down directly.[33] The series representations obtained resemble those encountered previously in the scalar problems [see Eqs. (2.11a) and (2.30a)], and have similar convergence properties. For plane wave scattering observed at large distances from the cone tip, the series converge very slowly so that termwise summation becomes impractical. As before, however, approximate closed-form expressions may be obtained for the backscattering from wide-angle and narrow-angle cones. For plane wave incidence the solution for the total electric field along the cone axis may be shown to be given by (Note: $\mathbf{x}_0=\boldsymbol{\theta}_0\cos\phi-\boldsymbol{\Phi}_0\sin\phi$)

$$r\mathbf{E}(r) = -\mathbf{x}_0(\pi/k)(S_1+S_2), \tag{3.6}$$

$$S_1 = \sum_p\left\{\frac{P_p^1(-\cos\theta_0)}{\sin p\pi(\partial/\partial p)P_p^1(\cos\theta_0)}\right\}(p+\tfrac{1}{2})e^{-i(p+1)\pi/2}\mathrm{j}_p'(kr);$$

$$P_p^1(\cos\theta_0)=0, \tag{3.6a}$$

and

$$S_2 = \frac{1}{i}\sum_s\left\{\frac{(d/d\theta_0)P_s^1(-\cos\theta_0)}{\sin s\pi(\partial^2/\partial s\partial\theta_0)P_s^1(\cos\theta_0)}\right\}(s+\tfrac{1}{2})$$

$$\times e^{-i(s+1)\pi/2}\mathrm{j}_s(kr); \quad \frac{d}{d\theta_0}P_s^1(\cos\theta_0)=0. \tag{3.6b}$$

In obtaining Eqs. (3.6), a factor $[i\zeta kJ\exp(ikr')]/4\pi r'$ has been set equal to unity to normalize the result to an incident wave of unit intensity (see Appendix B). In addition, a time dependence $\exp(-i\omega t)$ has been adopted by letting $j\rightarrow -i$ in conformity with the convention adopted in Section II. As in the scalar case it will be found convenient to replace the Bessel functions in Eqs. (3.6) by their integral form given in Eq. (2.14). For $kr\gg 1$, only the highest order term in j_ν' is retained. One obtains, then,

$$S_1+S_2 \cong \frac{1}{2\pi}\left(\frac{\pi kr}{2}\right)^{\frac{1}{2}}e^{i\pi/4}\int_W d\psi e^{ikr\cos\psi}$$

$$\times[S_1'(\psi)\cos\psi+S_2'(\psi)], \tag{3.7}$$

where

$$S_1'(\psi) = \sum_p\left\{\frac{P_p^1(-\cos\theta_0)}{\sin p\pi(\partial/\partial p)P_p^1(\cos\theta_0)}\right\}$$

$$\times(p+\tfrac{1}{2})e^{i(p+\frac{1}{2})(\psi-\pi)}, \tag{3.7a}$$

$$S_2'(\psi) = -\sum_s\left\{\frac{(d/d\theta_0)P_s^1(-\cos\theta_0)}{\sin s\pi(\partial^2/\partial s\partial\theta_0)P_s^1(\cos\theta_0)}\right\}$$

$$\times(s+\tfrac{1}{2})e^{i(s+\frac{1}{2})(\psi-\pi)}. \tag{3.7b}$$

The path of integration W in the ψ plane is shown in Fig. 2.

[30] Reference 10, Eqs. (A-1).
[31] Reference 9, Appendix 1.
[32] See J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 564.
[33] Reference 10, Eqs. (2.12).

B. Approximations for a Wide-Angle Cone: $\theta \approx \pi/2$

The large values of p and s may be obtained from the asymptotic expression given in Eq. (2.17). The expression for p is given in Eq. (2.18); for s one finds:

$$s \approx (n+\tfrac{3}{4})\alpha - \tfrac{1}{2}, \quad n = \text{large positive integer}, \tag{3.8}$$

where $\alpha = \pi/\theta_0$. It can be checked that the expressions inside the braces under the summation signs in Eqs. (3.7a) and (3.7b) evaluate to $(1/\theta_0)$ when the asymptotic form of the Legendre functions as well as that of the zeros is employed. As in the scalar case it can be verified that the asymptotic results above yield exact values over the entire range when $\theta_0 = \pi/2$. Therefore, it may again be assumed that the asymptotic approximations are approximately valid over the entire range when $\theta_0 \approx \pi/2$. When $\theta_0 = \pi/2$ the lowest pertinent values of p and s are 2 and 1, respectively, so that m and n in Eqs. (2.18) and (3.8) take on all integral values from 1 to ∞ and zero to ∞, respectively. If the approximate expressions above are substituted into Eqs. (3.7a) and (3.7b) one obtains series of the form encountered in Eq. (2.19a) which are summed readily. The result is

$$S_1' \approx -\frac{\alpha}{4\theta_0}\frac{5/4-(1/4)e^{i\alpha(\psi-\pi)}}{\sin^2[(\alpha/2)(\psi-\pi)]}e^{i(\alpha/4)(\psi-\pi)}, \tag{3.9a}$$

$$S_2' \approx \frac{\alpha}{4\theta_0}\frac{3/4+(1/4)e^{i\alpha(\psi-\pi)}}{\sin^2[(\alpha/2)(\psi-\pi)]}e^{-i(\alpha/4)(\psi-\pi)}. \tag{3.9b}$$

C. Approximations for a Narrow-Angle Cone: $\theta_0 \approx \pi$

When $\theta_0 \approx \pi$, the zeros p are given approximately by Eq. (2.20). In a manner similar to that employed to obtain the zeros p it can be shown that the zeros s may be expressed approximately as

$$s \approx n - \Delta_n, \quad n = 1, 2, 3 \cdots. \tag{3.10}$$

The approximations are valid if $\Delta_n \ll 1$. By employing the approximate expressions in Eqs. (2.20), (2.21), and (3.10) it is found that the series S_1' in Eq. (3.7a) is given in terms of the two series defined in Eq. (2.23) as

$$S_1' \approx (1/\pi)[S_1 + ((\pi-\theta_0)/2)^2 S_2]. \tag{3.11}$$

The justification of the approximations and mathematical manipulations here is analogous to that in Sec. IIA-3. Equation (3.11) may be manipulated into the same form as Eq. (2.24), except that the sum S_2 starts with $m=1$ instead of $m=0$, and the closed-form evaluation is carried out in an analogous fashion. The result may be written symbolically as follows:

$$S_1'(\psi) \approx M(\psi) - [(\pi-\theta_0)/2]^2 N(\psi). \tag{3.12}$$

Similarly, it is found that the series $-S_2'$ in Eq. (3.7b) is given by the same expression (3.12) except that the minus sign after $M(\psi)$ is changed to plus. Therefore, the factor in the integrand of Eq. (3.7) may be written as

$$S_1' \cos\psi + S_2' \approx (\cos\psi - 1)M - [(\pi-\theta_0)/2]^2(\cos\psi+1)N. \tag{3.13}$$

D. Asymptotic Evaluation

The asymptotic evaluation of the integral in Eq. (3.7) is carried out as in the scalar case. The final first-order result for the total electric field along the cone axis is found to be (for $kr \gg 1$) (see also reference 5)

$$\mathbf{E}(r) \sim \mathbf{x}_0[e^{-ikr} - e^{ikr}\sin^3\theta_0\{\ \}], \quad \xi \text{ small}, \tag{3.14a}$$

$$\sim \mathbf{x}_0\left[e^{-ikr} - \frac{i}{4(\theta_0-\pi/2)^2}\frac{e^{ikr}}{kr}\right],$$

$$\xi \gg 1, \quad \theta_0 \approx \pi/2, \tag{3.14b}$$

$$\sim \mathbf{x}_0\left[e^{-ikr} - i\left(\frac{\pi-\theta_0}{2}\right)^2\frac{e^{ikr}}{kr}\right], \quad kr \gg 1. \tag{3.14c}$$

The expression inside the braces in Eq. (3.14a) is identical with that in Eq. (2.26a). For $\theta_0 = \pi/2$, one obtains correctly the result of plane wave reflection from an infinite plane. Equations (3.14a) and (3.14b) are obtained by employing Eqs. (3.9), while Eq. (3.14c) arises from Eq. (3.13). In the latter case it is found that M has a double pole at $\psi = \pi$, N is $0(1/(\psi-\pi))$ at $\psi = \pi$, and neither M nor N has other poles in the range under consideration. Thus, it is seen that, to first order, only N contributes to the backscattered wave arising from saddle point at $\psi = 0$, and that M contributes only to the incident wave arising from the saddle point at $\psi = \pi$. This again serves as a check on the consistency of the approximations employed in the summation of the series.

If, for $\xi \ll 1$, the expression for $\mathbf{E}$ is written in the form of Eq. (2.12), it is seen that the function $f(k,\theta_0)$ is given by $(\pi/2k)[S_1'(0)+S_2'(0)]$, where S_1' and S_2' are defined in terms of infinite series in Eqs. (3.7a) and (3.7b). This same conclusion had been reached previously by Hansen and Schiff,[34] who employed for their analysis an approach similar to that described in the paragraph following Eq. (2.12). The remarks made after Eq. (2.27) apply to the convergence of these series as well. The direct summation of these divergent series as carried out to second order by Siegel *et al.*, yields the following result for the differential backscattering cross section[35]:

$$\sigma_D(0) = \frac{\lambda^2}{64\pi^2\cos^4\theta_0}(1 - 2\cos^2\theta_0 + \cdots), \quad \theta_0 \approx \pi/2 \tag{3.15a}$$

$$= \frac{\lambda^2\gamma^2}{4\pi^2}(1+6\gamma+\cdots), \quad \gamma = \sin^2\left(\frac{\pi-\theta_0}{2}\right),$$

$$\theta_0 \approx \pi. \tag{3.15b}$$

[34] Reference 15, Eq. (20).
[35] Reference 17, Eqs. (V-23) and (V-27).

It is verified directly that the values of $\sigma_D(0)$ computed from Eqs. (3.14) and (3.15) agree to first order.

It is of interest to note that the differential backscattering cross section derived by Spencer[11] from physical optics,

$$\sigma_D(0)=\frac{\lambda^2\tan^4(\pi-\theta_0)}{64\pi^2}, \tag{3.16}$$

agrees with that obtained for the vector diffraction problem in the limiting cases of wide-angle and narrow-angle cones. This result is remarkable since the physical optics approach assumes that the surface of the scatterer is smooth so that reflection at a point takes place as if the point were contained in an infinite plane tangent to the obstacle. It is apparent that this condition is definitely not satisfied at or near the vertex of a cone. The approximate validity of the physical optics formula in the range between large and small cone angles has been confirmed by the experimental measurements of Sletten.[12] It appears, therefore, that the physical optics expression furnishes at worst a good approximation to the backscattering cross section over the entire range $\pi/2<\theta_0\leqslant\pi$, $\xi=(2kr)^{\frac{1}{2}}|\cos\theta_0|\gg 1$.

A plot of the differential backscattering cross section obtained from Eqs. (3.14b) to (3.16) is shown in Fig. 6. The experimental results of Sletten are included as well. It is seen that as in the scalar case (see Fig. 4), the first order results shown in curves A and C are too large and too small, respectively. The second-order results of Siegel *et al.* (curves B and D), on the other hand, exhibit a common point at $\theta_0\approx 130°$ and have slopes at that point which are not markedly different so that they probably predict correctly the order of magnitude of $\sigma_D(0)$ over the entire range. It is interesting to note the proximity of these curves to the physical optics approximation E.

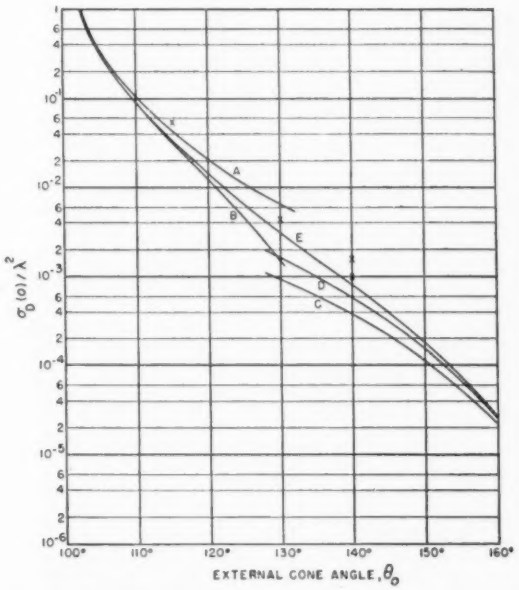


FIG. 6. Differential backscattering cross section (vector problem). A—1st order ($\theta_0\approx\pi/2$); B—2nd order ($\theta_0\approx\pi/2$) (Siegel *et al.*); C—1st order ($\theta_0\approx\pi$); D—2nd order ($\theta_0\approx\pi$) (Siegel *et al.*); E—physical optics (Spencer); ×—experimental points (Sletten).

APPENDIX A

Saddle Point Evaluation in the Vicinity of a Double Pole

The integral to be evaluated is

$$I=\int_{W_1} d\psi e^{ikr\cos\psi}S(\psi), \tag{A1}$$

where $kr\gg 1$ and W_1 is the steepest descent path through the saddle point at $\psi=0$ (see Fig. 2). The change of variable

$$\cos\psi=1+is^2, \quad -\infty<s<\infty, \tag{A2}$$

transforms the integral into

$$I=e^{ikr}\int_{-\infty}^{\infty} ds\exp(-krs^2)Q(s), \quad Q(s)=S(\psi)d\psi/ds. \tag{A3}$$

Let $Q(s)$ have a double pole at $s=b$, where b is small. In analogy to the approach employed by Van der Waerden[27] for the case when $Q(s)$ has a simple pole at $s=b$, the contribution at the double pole may be separated out as follows [assuming for the moment that the expansion contains no term in $1/(s-b)$]:

$$Q(s)=\beta/(s-b)^2+T(s). \tag{A4}$$

$T(s)$ is a function of s which has no singularities at $s=0$, b, and can therefore be expanded in a power series about $s=0$:

$$T(s)=t_0+t_1s+t_2s^2+\cdots. \tag{A5}$$

The contribution of $T(s)$ to I is readily evaluated. The remaining integral is

$$I'=\beta\int_{-\infty}^{\infty} ds\exp(-krs^2)/(s-b)^2. \tag{A6}$$

Upon employing the identity

$$\frac{\beta}{(s-b)^2}=\frac{\beta}{b^2}\left[\frac{2s^4}{(s^2-b^2)^2}-\frac{3s^2}{s^2-b^2}+\frac{2b^3s}{(s^2-b^2)^2}+1\right] \tag{A7}$$

in Eq. (A6) and integrating termwise one obtains

$$I'=(\beta/b^2)[4I_1-6I_2+(\pi/kr)^{\frac{1}{2}}], \tag{A8}$$

where

$$I_1=\int_0^{\infty}\frac{s^4\exp(-krs^2)}{(s^2-b^2)^2}ds; \quad I_2=\int_0^{\infty}\frac{s^2\exp(-krs^2)}{s^2-b^2}ds. \tag{A9}$$

The integral I_2 arises from a simple pole at $s=b$ and is

given by[27]

$$I_2=\tfrac{1}{2}(\pi/kr)^{\frac{1}{2}}-\tfrac{1}{2}(-\pi b^2)^{\frac{1}{2}}\exp(-krb^2)Q(-krb^2,\tfrac{1}{2}), \quad \text{(A10)}$$

where

$$Q(x,\tfrac{1}{2})=\int_x^\infty dx e^{-x}x^{-\frac{1}{2}} \quad \text{(A10a)}$$

is an incomplete gamma function which is tabulated.[36] For subsequent applications, b^2 is positive imaginary. In that case the sign of $(-b^2)^{\frac{1}{2}}$ is chosen such that[37]

$$(-b^2)^{\frac{1}{2}}=e^{-i\pi/4}|b|. \quad \text{(A11)}$$

To evaluate $I_1=I_1(kr)$, we successively multiply by $\exp(krb^2)$ and differentiate, with the assumption that $b^2<0$. Then, if $x=kr$,

$$\frac{d^2}{dx^2}[\exp(xb^2)I_1(x)]=\exp(xb^2)\int_0^\infty ds\, s^4\exp(-xs^2)$$

$$=\tfrac{1}{2}\exp(xb^2)\Gamma(5/2)x^{-5/2}. \quad \text{(A12)}$$

Thus, by repeated integration

$$I_1(x)=\frac{3\pi^{\frac{1}{2}}}{8}\exp(-xb^2)\int_x^\infty dz\int_z^\infty dy\exp(yb^2)y^{-5/2} \quad \text{(A13)}$$

$$=\frac{1}{2}\left(\frac{\pi}{x}\right)^{\frac{1}{2}}-\frac{b^2}{2}(\pi x)^{\frac{1}{2}}+(-\pi b^2)^{\frac{1}{2}}$$

$$\times\exp(-xb^2)\left(\frac{xb^2}{2}-\frac{3}{4}\right)Q(-xb^2,\tfrac{1}{2}). \quad \text{(A14)}$$

Since I_1 as given in Eq. (A9) is valid not only for $b^2<0$ but for all b^2 not positive real, and since Eq. (A14) is identical with Eq. (A9) along the negative real b^2 axis, the expression in Eq. (A14) may be continued into the complex b^2 domain. Upon substituting Eqs. (A10) and (A13) into (A8) and employing (A5) one obtains the asymptotic result for the integral in Eq. (A1):

$$I\sim-2\beta(\pi kr)^{\frac{1}{2}}e^{ikr}[1-e^{-i\pi/4}(kr)^{\frac{1}{2}}|b|\exp(-krb^2)$$

$$\times Q(-krb^2,\tfrac{1}{2})-[t_0/2\beta kr)]+\cdots]. \quad \text{(A15)}$$

When $(kr)^{\frac{1}{2}}|b|\gg1$, the double pole no longer lies near the saddle point. The incomplete gamma function in Eq. (A15) may then be reduced by integration by parts to yield the usual direct result

$$I\sim e^{ikr}(\pi/kr)^{\frac{1}{2}}[t_0+(\beta/b^2)]+\cdots. \quad \text{(A16)}$$

For the problems discussed in the text, the expansion for $Q(s)$ takes on the following form:

$$Q(s)=\beta/(s-b)^2+\beta'/(s-b)+t_0'+t_1s+\cdots. \quad \text{(A17)}$$

[36] Cf. J. Barkley Rosser, *On the Theory and Application of* $\int \exp(-x^2)dx$ (Mapleton House, Brooklyn, 1948).
[37] See H. Ott, Ann. Physik **43** (1943).

The contribution from the simple pole in this representation can be taken into account by Eq. (A10), after employing the identity

$$\frac{1}{s-b}=\frac{1}{b}\left(\frac{s^2+sb}{s^2-b^2}-1\right). \quad \text{(A18)}$$

For small values of $\xi=(2kr)^{\frac{1}{2}}|\cos\theta_0|$, the double pole contribution alone is of importance and leads to the results quoted in the text. For intermediate values of ξ the simple-pole contribution must be considered as well.

APPENDIX B

Normalizations for Unit Intensity Incident Plane Waves

(a) *Scalar Case*

When the point source moves to infinity, i.e., $r'\to\infty$, plane wave incidence obtains. For values of $r\ll r'$,

$$|\mathbf{r}-\mathbf{r}'|\to r'-r\cos\gamma, \quad \text{(B1)}$$

where γ is the angle between the vectors $\mathbf{r}$ and $\mathbf{r}'$. Thus,

$$\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}\to\left(\frac{e^{ikr'}}{4\pi r'}\right)e^{-ikr\cos\gamma}, \quad r'\to\infty. \quad \text{(B2)}$$

If the plane wave incident along the vector $\mathbf{r}'$ is to be of unit intensity, the factor inside the brackets in Eq. (B2) is set equal to unity.

(b) *Vector Case*

For an electric dipole of strength J in free space oriented on the polar axis in a spherical coordinate system as shown in Fig. 5a, the considerations of Eqs. (3.1)–(3.5) in the text apply directly. However, in view of the absence of the wave-guide walls, the scalar mode functions are those for free space.[7] It is then found that the total electric field on the polar axis is given for a plane wave by [letting $r'\to\infty$ and employing Eq. (2.13)]

$$r\mathbf{E}(\mathbf{r})=-\frac{J}{r'}\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}\frac{e^{ikr'}}{8\pi}\sum_{n=1}^{\infty}(-i)^{n+1}(2n+1)$$

$$\times[i\mathrm{j}_n'(kr)+\mathrm{j}_n(kr)]\mathbf{x}_0 \quad \text{(B3)}$$

$$=\left[J\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}ik\frac{e^{ikr'}}{4\pi r'}\right]e^{-ikr}\mathbf{x}_0, \quad \text{(B4)}$$

where use has been made of the formula[38]

$$kre^{ikr}=\sum_{n=0}^{\infty}(2n+1)i^n\mathrm{j}_n(kr). \quad \text{(B5)}$$

[38] See reference 2, p. 144.

For an incident wave of unit intensity, the expression inside the square brackets in Eq. (B4) is set equal to unity.

ACKNOWLEDGMENT

The research described in this paper was sponsored in part by the Air Force Cambridge Research Center under Contract No. AF-19(604)-890.

This paper comprises part of a doctoral dissertation entitled "Diffraction by Wedges and Cones" submitted to the Graduate Faculty of the Polytechnic Institute of Brooklyn in June, 1952, in partial fulfilment of the requirements for the degree of Doctor of Electrical Engineering. The author expresses his sincere appreciation to his advisor, Professor N. Marcuvitz, for his valuable advice and criticism.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 2 FEBRUARY, 1955

Diffraction by Apertures

CHAANG HUANG, *Sylvania Electric Products Company, Ipswich, Massachusetts*

AND

RALPH D. KODIS AND HAROLD LEVINE, *Cruft Laboratory, Harvard University, Cambridge, Massachusetts*

(Received February 27, 1954)

The diffraction of plane electromagnetic waves by apertures in a plane screen which is infinitesimally thin and perfectly conducting is studied both theoretically and experimentally. The theoretical analysis employs a dyadic Green's function to develop vector formulas for the scattered fields, and from these formulas integral equations are obtained for the aperture distributions. The vector representation makes possible a compact demonstration of the electromagnetic form of Babinet's principle by means of which one may extend the aperture analysis to complementary disks. The integral equations are then used to construct a variational principle for the aperture transmission coefficient.

Detailed analysis and numerical computations are carried out for two configurations. For the circular aperture a first-order vector trial function with frequency dependent coefficients is chosen for the aperture distribution. The approximate transmission coefficient is found to agree closely with the exact value in the region $2\pi a/\lambda \leq 3$. For elliptical apertures a zeroth-order approximation is evaluated using a one-component trial function. Numerical results are given for minor-to-major axis ratios of $\frac{1}{2}$ and $\frac{1}{3}$.

Transmission coefficient measurements were carried out in the 24 000 megacycle band ($\lambda = 1.25$ cm) using an image plane technique. The apparatus was first calibrated with the exact solution of the circular aperture. The approximate results calculated for elliptical apertures are then seen to be in good agreement with the measurements over the accessible range.

1. INTRODUCTION

THE problem of the diffraction of plane electromagnetic waves by apertures in a conducting plane screen has already been considered by many authors in both its theoretical and experimental aspects. It is the purpose of the work reported in this paper to supplement the earlier results in two ways: first, by improving the usual "classical" approximations through a systematic theoretical procedure that is inherently more precise; second, by showing that measurements at microwave frequencies can determine the degree of improvement quite accurately.

The difficulties which attend any analysis of the aperture problem are made plain by a study of some of the methods that have been used. The most direct approach, from one point of view, is to find a series expansion for the field in terms of suitable characteristic functions. Although a few of the simplest aperture diffraction problems can be solved this way in principle, the method becomes extremely complicated when the solution must satisfy a *vector* wave equation. Such an expansion has been carried out in detail only for a circular aperture and was first attempted by Möglich for the complementary problem of the circular disk. In the procedure, as he developed it, the rectangular components of the Hertz vectors that describe the incident and scattered fields are expanded in series of appropriate oblate spheroidal wave functions. The coefficients in these expansions are determined by the boundary conditions at the surface of the disk and the "edge condition." The meaning of the second requirement was made clear by Meixner[2] and Andrejewski[3] who pointed out that the edge condition is necessary for uniqueness and is equivalent to the statement that the energy density must be finite in the neighborhood of an edge where the normal component of the electric field is singular. With explicit series for the Hertz vector all other quantities of interest may be determined, and Babinet's principle makes it easy to relate the results to the complementary problem of a circular aperture.

A low-frequency solution for the circular disk has

[1] F. Möglich, Ann. Physik [4] **83**, 609 (1927).
[2] J. Meixner, Natur **3A**, 506 (1948); Ann. Physik [6] **6**, 1 (1949).
[3] J. Meixner and W. Andrejewski, Ann. Physik [6] **7**, 157 (1950).

been effected by Bouwkamp[4] who obtains the first few terms of the series expansion for the fields in powers of ka. His method proceeds by transforming Maxwell's equations for the fields into integral equations which incorporate the boundary conditions. The solution of these integral equations is then found by determining the coefficients of power series expansions for the unknown functions.

Although these exact methods clear up many delicate points connected with the type of diffraction problem being considered here, they are not particularly helpful in the analysis of scattering by apertures of more general shape. However, a number of formulations have long been available for the analysis of scattering by such apertures. To the extent that they are solutions of Maxwell's equations all these formulations are equivalent; but when explicit results are desired for a particular case, the necessary approximations give differing degrees of accuracy depending upon the frequency range that is of interest and the formulation that is used. Approximations which have proved useful may be classified as follows:

1. *The low-frequency (static) approximation for small apertures*
   a. Retardation is neglected.
   b. The incident field vectors are assumed to be constant over the aperture.
2. *The high-frequency (optical) approximation for large apertures*
   a. The tangential component of the magnetic field is assumed to vanish on the dark side of the conducting screen.
   b. The tangential component of the electric field in the aperture is assumed to be equal to that of the incident electric field.

The static approximations have been studied by Rayleigh,[5] Bethe,[6] Bouwkamp,[4] and Tai.[7] Their results are accurate only for frequencies low enough so that $ka=2\pi a/\lambda \ll 1$ (here a is a characteristic dimension of the aperture or disk). The optical approximations, which are of the Kirchhoff type, have been discussed by Stratton and Chu,[8] Schelkunoff,[9] Silver and Ehrlich,[10] and Levine and Schwinger.[11] Among these formulations the analysis of Levine and Schwinger has the advantage of a relative analytic simplicity, and its application to annular, elliptic, and rectangular apertures will be discussed in a future paper.

In contrast to the above methods, Levine and Schwinger have derived vector integral equations for the field in the aperture, or on the plane screen, and they have shown how to calculate the far-zone diffracted fields and the transmission coefficients of apertures in terms of variational principles related to these integral equations. Thus, by choosing an appropriate trial function for either the tangential component of the electric field in the aperture or for the magnetic field on the back side of the diffracting screen, an accurate result can be obtained for the transmission coefficient without making any further assumptions. In this way they have computed the transmission coefficient of a circular aperture for normal incidence using two different trial functions for the tangential component of the electric field in the aperture. The first of these functions is constructed from a single component electric field with no angular variation and with a radial dependence corresponding to that of the static solution. It results in a transmission coefficient that is accurate only at low frequencies corresponding to $ka<1$. The second trial function is the exact static aperture field obtained by Bouwkamp. It gives an improved result that is still accurate at frequencies up to $ka=2$ for which the Bouwkamp field itself is no longer the correct aperture distribution.

Since these trial functions are frequency independent they give unsatisfactory answers when $ka>2$. In this study a frequency dependent trial function is constructed for the circular aperture, and it will be seen that a very high degree of accuracy over a wide range of frequencies can be achieved with only moderate analytic complexity. Finally, the variational procedure is applied to elliptic apertures using a single component trial function analogous to the first one used by Levine and Schwinger for the circular aperture.

The analytic difficulties encountered in problems of diffraction by apertures and disks make it easier in many cases to obtain the quantities of physical interest by direct experimental measurement. For example, Andrews[12] and Silver and Ehrlich[10] have made detailed measurements in the near zone of diffracting apertures and disks. Their results are in agreement with many theoretical predictions, such as the behavior of the tangential component of the magnetic field in the aperture and the axial distribution of the field vectors. In the far zone, Aden[13] and Sevick[14] have measured the backscattering from spheres and coupled antennas by methods which can be extended easily to disks.

It seems, however, that the transmission coefficients of apertures have never been measured. In this research, their measurement is carried out at K-band with equipment which was originally built and used by Kodis[15] for determining the near-zone diffraction pat-

[4] C. J. Bouwkamp, Philips Research Reports No. 5, 401–442 (December, 1950).
[5] Lord Rayleigh, Phil. Mag. **44**, 28 (1897).
[6] H. A. Bethe, Phys. Rev. **66**, 163 (1944).
[7] C. T. Tai, Trans. Inst. Radio Engrs. (February 1952).
[8] J. A. Stratton and L. J. Chu, Phys. Rev. **56**, 99 (1939).
[9] S. A. Schelkunoff, Comm. on Pure and Appl. Math. **4** (1) (June 1951).
[10] S. Silver, Report No. 163, Antenna Laboratory, University of California, Berkeley (1949); M. J. Ehrlich, Doctoral thesis, University of California (1951).
[11] H. Levine and J. Schwinger, Comm. on Pure and Appl. Math. **3**, (4) (December 1950).
[12] C. L. Andrews, J. Appl. Phys. **21**, 761 (1950).
[13] A. L. Aden, J. Appl. Phys. **22**, 601 (1951).
[14] J. Sevick, Ph.D. thesis, Harvard University (June 1952).
[15] R. D. Kodis, J. Appl. Phys. **23**, 249 (1952); Cruft Laboratory Technical Report No. 105 (June 1950).

terns of cylinders. This equipment has been adapted to the measurement of far-zone diffracted amplitudes, and after being calibrated with circular apertures is used to check the range of validity of the approximate theories applied to more complicated aperture shapes for which no exact results are available.

2. THE INTEGRAL EQUATIONS OF DIFFRACTION THEORY

Schwinger and Levine have shown that the electromagnetic fields in a source-free region can be represented by integral formulas involving appropriate dyadic Green's functions and the values of the tangential component of either the electric or magnetic vectors on the surface bounding the region. If the surface is denoted by S, the integrals are

$$\left.\begin{aligned}\mathbf{E}(\mathbf{r})&=-\int_S dS'[\mathbf{n}'\times\mathbf{E}(\mathbf{r}')]\cdot[\nabla'\times\boldsymbol{\Gamma}^{(1)}(\mathbf{r}',\mathbf{r})]\\ \mathbf{H}(\mathbf{r})&=ik\int_S dS'[\mathbf{n}'\times\mathbf{E}(\mathbf{r}')]\cdot\boldsymbol{\Gamma}^{(2)}(\mathbf{r}',\mathbf{r})\end{aligned}\right\} \quad (2\text{-}1)$$

or

$$\left.\begin{aligned}\mathbf{E}(\mathbf{r})&=-ik\int_S dS'[\mathbf{n}'\times\mathbf{H}(\mathbf{r}')]\cdot\boldsymbol{\Gamma}^{(2)}(\mathbf{r}',\mathbf{r})\\ \mathbf{H}(\mathbf{r})&=-\int_S dS'[\mathbf{n}'\times\mathbf{H}(\mathbf{r}')]\cdot[\nabla'\times\boldsymbol{\Gamma}^{(1)}(\mathbf{r}',\mathbf{r})]\end{aligned}\right\}, \quad (2\text{-}2)$$

where $\mathbf{n}'$ is the unit outward normal at S. The functions $\boldsymbol{\Gamma}^{(1)}$ and $\boldsymbol{\Gamma}^{(2)}$ are called the electric and magnetic dyadic Green's functions. They may be interpreted as operators which transform a unit electric or magnetic current vector at $\mathbf{r}'$ in a bounded region into an electric or magnetic field vector at $\mathbf{r}$ in the same region. Their functional representation depends only on the geometric shape of the surface S, and they have the following mathematical properties

$$\left.\begin{aligned}\nabla\times\nabla\times\boldsymbol{\Gamma}(\mathbf{r}',\mathbf{r})-k^2\boldsymbol{\Gamma}(\mathbf{r}',\mathbf{r})&=\boldsymbol{\varepsilon}\delta(\mathbf{r}'-\mathbf{r})\\ \left.\begin{aligned}\mathbf{n}'\times\boldsymbol{\Gamma}^{(1)}(\mathbf{r}',\mathbf{r})&=0\\ \mathbf{n}'\times[\nabla\times\boldsymbol{\Gamma}^{(2)}(\mathbf{r}',\mathbf{r})]&=0\end{aligned}\right\}&\ \mathbf{r}\text{ on }S\\ \boldsymbol{\Gamma}(\mathbf{r},\mathbf{r}')&=[\boldsymbol{\Gamma}(\mathbf{r}',\mathbf{r})]^T\\ \nabla\times\boldsymbol{\Gamma}^{(1)}(\mathbf{r},\mathbf{r}')&=[\nabla'\times\boldsymbol{\Gamma}^{(2)}(\mathbf{r}',\mathbf{r})]^T\end{aligned}\right\}. \quad (2\text{-}3)$$

Here, $\boldsymbol{\varepsilon}$ represents the unit dyadic, $\delta(\mathbf{r}'-\mathbf{r})$ is the Dirac delta function, and $\boldsymbol{\Gamma}^T$ denotes the transposed dyadic of $\boldsymbol{\Gamma}$.

If the region containing the perfectly conducting surfaces is unbounded, the formulas for the diffracted fields require only the free-space dyadic Green's function $\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\mathbf{r}')$. Then

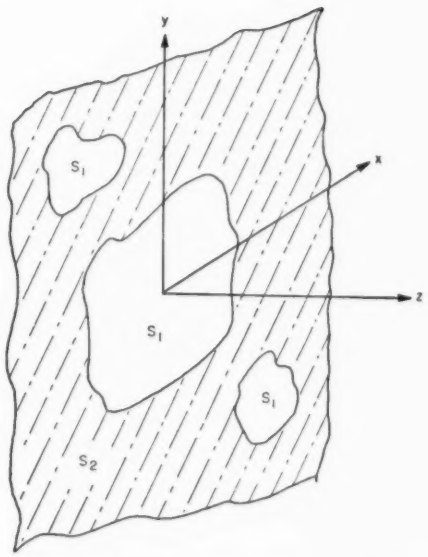


FIG. 1. Diffracting apertures in a plane screen.

$$\left.\begin{aligned}\mathbf{E}(\mathbf{r})&=-ik\int_S dS'[\mathbf{n}'\times\mathbf{H}(\mathbf{r}')]\cdot\boldsymbol{\Gamma}^{(0)}(\mathbf{r}',\mathbf{r})\\ \mathbf{H}(\mathbf{r})&=-\int_S dS'[\mathbf{n}'\times\mathbf{H}(\mathbf{r}')]\cdot[\nabla'\times\boldsymbol{\Gamma}^{(0)}(\mathbf{r}',\mathbf{r})]\end{aligned}\right\}, \quad (2\text{-}4)$$

where

$$\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\mathbf{r}')=\left(\boldsymbol{\varepsilon}-\frac{1}{k^2}\nabla\nabla'\right)\frac{\exp(ik|\mathbf{r}-\mathbf{r}'|)}{4\pi|\mathbf{r}-\mathbf{r}'|}$$
$$=\boldsymbol{\Gamma}^{(0)}(\mathbf{r}',\mathbf{r}). \quad (2\text{-}5)$$

It should be noted that in this case one part of the surface of integration is the surface of a very large sphere.

To apply the integral formulas to the diffraction of electromagnetic waves by a perforated plane screen, it is convenient to orient the screen in the plane $z=0$ (Fig. 1). The half-spaces $z>0$ and $z<0$ are then considered separately. The surface of integration includes the perfectly conducting screen S_s on which $\mathbf{e}_z\times\mathbf{E}=\mathbf{e}_z\cdot\mathbf{H}=0$, the aperture S_a, and the surface of a large hemisphere centered on the aperture. If the dyadic Green's function for each half-space $z\lesseqgtr 0$ is denoted by $\boldsymbol{\Gamma}_{\mp}$, then (2-1) can be reduced to

$$\left.\begin{aligned}\mathbf{E}(\mathbf{r})&=\int_{S_a}[\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\rho}')]\cdot[\nabla'\times\boldsymbol{\Gamma}_+^{(1)}(\boldsymbol{\rho}',\mathbf{r})]dS'\\ \mathbf{H}(\mathbf{r})&=-ik\int_{S_a}[\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\rho}')]\cdot\boldsymbol{\Gamma}_+^{(2)}(\boldsymbol{\rho}',\mathbf{r})dS'\end{aligned}\right\} z\gtreqless 0 \quad (2\text{-}6)$$

$$\left.\begin{aligned}\mathbf{E}(\mathbf{r})&=\mathbf{E}^{\text{inc}}(\mathbf{r})+\mathbf{E}^{\text{ref}}(\mathbf{r})\\&\quad-\int_{S_a}[\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]\cdot[\nabla'\times\boldsymbol{\Gamma}_-^{(1)}(\boldsymbol{\varrho}',\mathbf{r})]dS'\\\mathbf{H}(\mathbf{r})&=\mathbf{H}^{\text{inc}}(\mathbf{r})+\mathbf{H}^{\text{ref}}(\mathbf{r})\\&\quad+ik\int_{S_a}[\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]\cdot\boldsymbol{\Gamma}_-^{(2)}(\boldsymbol{\varrho}',\mathbf{r})dS'\end{aligned}\right\}z\leqq 0. \quad (2\text{-}7)^*$$

In obtaining these formulas use has been made of the boundary condition on the screen and of the radiation condition at infinity.

The dyadic Green's functions involved in Eqs. (2-6) and (2-7) have been shown to be related to the free-space dyadic Green's function as follows[11]:

$$\left.\begin{aligned}\boldsymbol{\Gamma}_+^{(1)}(\mathbf{r},\mathbf{r}')&=\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\mathbf{r}')-\boldsymbol{\Gamma}^{(0)}[\mathbf{r},\mathbf{r}'-2\mathbf{e}_z(\mathbf{e}_z\cdot\mathbf{r}')]\\&\quad\cdot(\boldsymbol{\varepsilon}-2\mathbf{e}_z\mathbf{e}_z),\quad z,z'\geqq 0\\\boldsymbol{\Gamma}_+^{(2)}(\mathbf{r},\mathbf{r}')&=\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\mathbf{r}')+\boldsymbol{\Gamma}^{(0)}[\mathbf{r},\mathbf{r}'-2\mathbf{e}_z(\mathbf{e}_z\cdot\mathbf{r}')]\\&\quad\cdot(\boldsymbol{\varepsilon}-2\mathbf{e}_z\mathbf{e}_z),\quad z,z'\geqq 0\end{aligned}\right\}, \quad (2\text{-}8)$$

and

$$\boldsymbol{\Gamma}_-(\mathbf{r},\mathbf{r}')=\boldsymbol{\Gamma}_+[\mathbf{r}-2\mathbf{e}_z(\mathbf{e}_z\cdot\mathbf{r}),\ \mathbf{r}'-2\mathbf{e}_z(\mathbf{e}_z\cdot\mathbf{r}')],\quad z,z'\leqq 0. \quad (2\text{-}9)$$

If Eq. (2-8) is post multiplied by the vector $\mathbf{e}_z\times\mathbf{E}(\mathbf{r}')$ and the result evaluated at $z'=0$, it can be seen that

$$\boldsymbol{\Gamma}_+^{(2)}(\mathbf{r},\boldsymbol{\varrho}')\cdot[\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]=2\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\boldsymbol{\varrho}')\cdot[\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]. \quad (2\text{-}10)$$

With this identity and the symmetry properties Eq. (2-3) between the dyads and their transposes, it is possible to rewrite Eqs. (2-6) and (2-7) in the simpler form

$$\left.\begin{aligned}\mathbf{E}(\mathbf{r})&=\nabla\times\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\boldsymbol{\varrho}')\cdot[2\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]dS'\\\mathbf{H}(\mathbf{r})&=-ik\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\boldsymbol{\varrho}')\cdot[2\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]dS'\end{aligned}\right\}z\geqq 0 \quad (2\text{-}11)$$

$$\left.\begin{aligned}\mathbf{E}(\mathbf{r})&=\mathbf{E}^{\text{inc}}(\mathbf{r})+\mathbf{E}^{\text{ref}}(\mathbf{r})\\&\quad-\nabla\times\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\boldsymbol{\varrho}')\cdot[2\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]dS'\\\mathbf{H}(\mathbf{r})&=\mathbf{H}^{\text{inc}}(\mathbf{r})+\mathbf{H}^{\text{ref}}(\mathbf{r})\\&\quad+ik\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\boldsymbol{\varrho}')\cdot[2\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]dS'\end{aligned}\right\}z\leqq 0. \quad (2\text{-}12)$$

* The superscripts inc and ref indicate the incident and specularly reflected waves from a plane conducting screen which is not perforated.

These formulas require knowledge of the tangential electric field in the aperture, which is not available in general.[16] It is possible, however, to formulate integral equations for the unknown function by making use of the continuity of either the normal component of the electric field or the tangential component of the magnetic field in the aperture. If the appropriate components of Eqs. (2-11) and (2-12) are set equal to each other in the aperture ($z=0$), then by noting the fact that in the plane of the aperture

$$\mathbf{e}_z\cdot\mathbf{E}^{\text{inc}}(\boldsymbol{\varrho})=\mathbf{e}_z\cdot\mathbf{E}^{\text{ref}}(\boldsymbol{\varrho})\quad\text{and}\quad\mathbf{e}_z\times\mathbf{H}^{\text{inc}}(\boldsymbol{\varrho})=\mathbf{e}_z\times\mathbf{H}^{\text{ref}}(\boldsymbol{\varrho}),$$

we can write the desired integral equations†

$$\left.\begin{aligned}\mathbf{e}_z\cdot\mathbf{E}^{\text{inc}}(\boldsymbol{\varrho})&=\mathbf{e}_z\cdot\nabla\times\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\\&\quad\cdot[2\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]dS'\\\mathbf{e}_z\times\mathbf{H}^{\text{inc}}(\boldsymbol{\varrho})&=-ik\mathbf{e}_z\times\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\\&\quad\cdot[2\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]dS'\end{aligned}\right\}\boldsymbol{\varrho},\ \boldsymbol{\varrho}'\text{ in }S_a. \quad (2\text{-}13)$$

An analogous procedure will yield a pair of integral equations for the surface current density on the screen. In this case the perforated plane conducting screen is considered to be an obstacle imbedded in free space, and Eq. (2-4) is used to describe the fields. With the radiation condition, the symmetry property of Eq. (2-3), and the notation

$$\mathbf{K}(\boldsymbol{\varrho})=\mathbf{e}_z\times[\mathbf{H}_-(\boldsymbol{\varrho})-\mathbf{H}_+(\boldsymbol{\varrho})],\quad\boldsymbol{\varrho}\text{ in }S_a, \quad (2\text{-}14)$$

(the subscripts + and − here denote the limiting values of $\mathbf{H}$ as $z\to 0$ from the positive and negative sides of the screen respectively), it is found that

$$\left.\begin{aligned}\mathbf{E}(\mathbf{r})&=\mathbf{E}^{\text{inc}}(\mathbf{r})-ik\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\boldsymbol{\varrho}')\cdot\mathbf{K}(\boldsymbol{\varrho}')dS'\\\mathbf{H}(\mathbf{r})&=\mathbf{H}^{\text{inc}}(\mathbf{r})-\nabla\times\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\boldsymbol{\varrho}')\cdot\mathbf{K}(\boldsymbol{\varrho}')dS'\end{aligned}\right\}. \quad (2.15)$$

[16] C. Huang, Cruft Laboratory Technical Report No. 163 (February 1953).

† It is worth pointing out in passing that formula (2-11) for the electric field in the half-space $z\geq 0$, when particularized to the normal component of that field in the aperture, gives

$$\mathbf{e}_z\cdot\mathbf{E}(\boldsymbol{\varrho})=\mathbf{e}_z\cdot\nabla\times\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\cdot[2\mathbf{e}_z\times\mathbf{E}(\boldsymbol{\varrho}')]dS'.$$

Comparison with the integral Eq. (2-13) shows that *in the aperture* $\mathbf{e}_z\cdot\mathbf{E}(\boldsymbol{\varrho})=\mathbf{e}_z\cdot\mathbf{E}^{\text{inc}}(\boldsymbol{\varrho})$. A similar argument can be made for tangential H. An alternative proof of these important boundary conditions appeals to symmetry arguments rather than to the integral equations. It can be found in references 11 and 16.

On the screen, $\mathbf{e}_z\times\mathbf{E}(\varrho)=\mathbf{e}_z\cdot\mathbf{H}(\varrho)=0$, so that

$$\left.\begin{aligned}\mathbf{e}_z\times\mathbf{E}^{\text{inc}}(\varrho)&=ik\mathbf{e}_z\times\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\varrho,\varrho')\cdot\mathbf{K}(\varrho')dS'\\ \mathbf{e}_z\cdot\mathbf{H}^{\text{inc}}(\varrho)&=\mathbf{e}_z\cdot\nabla\times\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\varrho,\varrho')\cdot\mathbf{K}(\varrho')dS'\end{aligned}\right\}\varrho\text{ in }S_a. \quad (2\text{-}16)$$

These are the required integral equations for $\mathbf{K}$.

3. BABINET'S PRINCIPLE

The integral equations given above make it possible to formulate directly Babinet's principle for the diffraction of electromagnetic waves by a perforated plane conducting screen. In optics the principle states simply that for Fraunhofer diffraction the patterns (light intensities) produced by complementary screens are the same. With electromagnetic waves, however, not only must the screens be complementary but the fields as well. Thus, in order to formulate an analogous electromagnetic principle a suitable complementary field must be defined which satisfies Maxwell's equations and leaves the Poynting vector unaltered (except for a possible constant). In terms of the original vectors, one possible set of new vectors (subscript 2) is

$$\mathbf{H}_2=\mathbf{E}_1,\quad \mathbf{E}_2=-\mathbf{H}_1 \text{ (Gaussian units)}.$$

The complementary field so defined is considered to be diffracted by thin coplanar conducting obstacles of area S_1; the original field is diffracted by apertures occupying the same regions S_1 in a conducting screen.

Under these circumstances the relationship between the two different waves may be extracted directly from the integral formulas of the previous section. With the appropriate subscripts Eq. (2-16) becomes

$$\left.\begin{aligned}\mathbf{e}_z\times\mathbf{E}_2^{\text{inc}}(\varrho)&=ik\mathbf{e}_z\times\int_{S_1}\boldsymbol{\Gamma}^{(0)}(\varrho,\varrho')\cdot\mathbf{K}_2(\varrho')dS'\\ \mathbf{e}_z\cdot\mathbf{H}_2^{\text{inc}}(\varrho)&=\mathbf{e}_z\cdot\nabla\times\int_{S_1}\boldsymbol{\Gamma}^{(0)}(\varrho,\varrho')\cdot\mathbf{K}_2(\varrho')dS'\end{aligned}\right\}. \quad (3\text{-}1)$$

In the complementary problem Eqs. (2-13) apply, and we have

$$\left.\begin{aligned}\mathbf{e}_z\cdot\mathbf{E}_1^{\text{inc}}(\varrho)&=\mathbf{e}_z\cdot\nabla\times\int_{S_1}\boldsymbol{\Gamma}^{(0)}(\varrho,\varrho')\cdot[2\mathbf{e}_z\times\mathbf{E}_1(\varrho')]dS'\\ \mathbf{e}_z\times\mathbf{H}_1^{\text{inc}}(\varrho)&=-ik\mathbf{e}_z\times\int_{S_1}\boldsymbol{\Gamma}^{(0)}(\varrho,\varrho')\cdot[2\mathbf{e}_z\times\mathbf{E}_1(\varrho')]dS'\end{aligned}\right\}. \quad (3\text{-}2)$$

If the complementary fields are substituted on the left of Eq. (3-2), comparison with Eq. (3-1) shows at once that

$$2\mathbf{e}_z\times\mathbf{E}_1(\varrho)=\mathbf{K}_2(\varrho). \quad (3\text{-}3)$$

This relation between the tangential component of the electric field in the apertures and the surface current on the complementary conductors makes it possible to write integral formulas for the diffracted fields, $\mathbf{E}_1$, $\mathbf{H}_1$ and $\mathbf{E}_2$, $\mathbf{H}_2$, in terms of $\mathbf{K}_2$ alone. For the apertures we make use of Eqs. (2-11) and (2-12) and find

$$\left.\begin{aligned}\mathbf{E}_1(\mathbf{r})&=\nabla\times\int_{S_1}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\varrho')\cdot\mathbf{K}_2(\varrho')dS'\\ \mathbf{H}_1(\mathbf{r})&=-ik\int_{S_1}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\varrho')\cdot\mathbf{K}_2(\varrho')dS'\end{aligned}\right\}z\geq 0 \quad (3\text{-}4)$$

$$\left.\begin{aligned}\mathbf{E}_1(\mathbf{r})&=\mathbf{E}_1^{\text{inc}}(\mathbf{r})+\mathbf{E}_1^{\text{ref}}(\mathbf{r})-\nabla\times\int_{S_1}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\varrho')\cdot\mathbf{K}_2(\varrho')dS'\\ \mathbf{H}_1(\mathbf{r})&=\mathbf{H}_1^{\text{inc}}(\mathbf{r})+\mathbf{H}_1^{\text{ref}}(\mathbf{r})+ik\int_{S_1}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\varrho')\cdot\mathbf{K}_2(\varrho')dS'\end{aligned}\right\}z\leq 0. \quad (3\text{-}5)$$

For obstacles in the complementary field Eq. (2-15) leads to

$$\left.\begin{aligned}\mathbf{E}_2(\mathbf{r})&=-\mathbf{H}_1^{\text{inc}}(\mathbf{r})-ik\int_{S_1}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\varrho')\cdot\mathbf{K}_2(\varrho')dS'\\ \mathbf{H}_2(\mathbf{r})&=\mathbf{E}_1^{\text{inc}}(\mathbf{r})-\nabla\times\int_{S_1}\boldsymbol{\Gamma}^{(0)}(\mathbf{r},\varrho')\cdot\mathbf{K}_2(\varrho')dS'\end{aligned}\right\}. \quad (3\text{-}6)$$

The electromagnetic form of Babinet's principle results when Eqs. (3-4), (3-5), and (3-6) are combined in such a way as to eliminate the integrals. It is

$$\left.\begin{aligned}\mathbf{E}_1-\mathbf{H}_2&=\mathbf{E}_1^{\text{ref}}=\mathbf{H}_2^{\text{ref}}\\ \mathbf{H}_1+\mathbf{E}_2&=\mathbf{H}_1^{\text{ref}}=-\mathbf{H}_2^{\text{ref}}\end{aligned}\right\}z\leq 0$$

$$\left.\begin{aligned}\mathbf{E}_1+\mathbf{H}_2&=\mathbf{E}_1^{\text{inc}}=\mathbf{H}_2^{\text{inc}}\\ \mathbf{H}_1-\mathbf{E}_2&=\mathbf{H}_1^{\text{inc}}=-\mathbf{E}_2^{\text{inc}}\end{aligned}\right\}z\geq 0.$$

So the solution of either plane diffraction problem is seen to imply the solution of its Babinet complement since the incident and specularly reflected waves are known.

4. VARIATIONAL APPROXIMATION FOR APERTURE TRANSMISSION COEFFICIENTS

Associated with the integral equations of Section II are a pair of variational principles for the far zone amplitude of the field transmitted through an aperture in a plane screen. The development of these principles has been carried out in detail by Levine and Schwinger who have also shown that they are directly related to the aperture transmission coefficient. The formulation

that involves the tangential component of the electric field in the aperture proceeds from Eq. (2-11), with the unit vector subscript $\mathbf{n}''$ attached to the field vector to denote the (arbitrary) direction of propagation of the incident wave. It can be shown that in $z>0$,

$$\mathbf{E}_{\mathbf{n}''}(\mathbf{r})=2\int_{S_a}[\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho}')]\cdot[\nabla'\times\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho}',\mathbf{r})]dS'$$

$$=2\int_{S_a}[\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho}')]\times\nabla'G(\boldsymbol{\varrho}',\mathbf{r})dS'.$$

By making use of the asymptotic form of the scalar Green's function,

$$G(\mathbf{r}',\mathbf{r})\simeq\frac{\exp[ik(r-\mathbf{n}\cdot\mathbf{r}')]}{4\pi r},\quad r\to\infty,$$

it is possible to write

$$\mathbf{E}_{\mathbf{n}''}(\mathbf{r})\simeq\mathbf{n}\times\mathbf{A}(\mathbf{n},\mathbf{n}'')\frac{\exp(ikr)}{r},\quad r\to\infty.$$

The amplitude $\mathbf{A}$ is a function of both the observation direction $\mathbf{n}$ and the incident direction $\mathbf{n}''$. It is given by the integral

$$\mathbf{A}(\mathbf{n},\mathbf{n}'')=\frac{ik}{2\pi}\int_{S_a}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho}')\exp(-ik\mathbf{n}\cdot\boldsymbol{\varrho}')dS'. \quad (4\text{-}1)$$

When the directions of incidence and observation coincide, the amplitude is related directly to the transmission coefficient. If this number is defined as the power transmitted through an aperture in the direction of incidence per unit incident power and per unit area of the aperture, it can be verified that (see Sec. 7)

$$t(\mathbf{n}'')=\frac{2\pi}{kS_a}Im\mathbf{h}\cdot\mathbf{A}(\mathbf{n}'',\mathbf{n}''), \quad (4\text{-}2)$$

where $\mathbf{h}$ is the unit magnetic polarization vector of the incident wave, and S_a is the aperture area.

In order to evaluate a variational approximation for $t(\mathbf{n}'')$ it is necessary to find a stationary expression for $\mathbf{h}\cdot\mathbf{A}(\mathbf{n}'',\mathbf{n}'')$. This can be constructed from the integral equation for the aperture field, Eq. (2-13), which is

$$\mathbf{e}_z\times\mathbf{h}\exp(ik\mathbf{n}''\cdot\boldsymbol{\varrho})=-2ik\mathbf{e}_z\times\int_{S_a}\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\cdot[\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho}')]dS',\quad \boldsymbol{\varrho}\text{ in }S_a.$$

Vector premultiplication of this equation by $\mathbf{E}_{\mathbf{n}}(\boldsymbol{\varrho})$ and integration over the aperture domain gives

$$\int_{S_a}[\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}}(\boldsymbol{\varrho})]\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\cdot[\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho}')]dSdS'$$

$$=\frac{i}{2k}\mathbf{h}\cdot\int_{S_a}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}}(\boldsymbol{\varrho})\exp(ik\mathbf{n}''\cdot\boldsymbol{\varrho})dS$$

$$=\frac{\pi}{k^2}\mathbf{h}\cdot\mathbf{A}(-\mathbf{n}'',\mathbf{n}). \quad (4\text{-}3)$$

By making use of the evident symmetry in the first term of Eq. (4-3) as regards interchange of the indices $\mathbf{n}$ and $\mathbf{n}''$, the usual reciprocity relation can be obtained,

$$\begin{aligned}&\mathbf{h}''\cdot\mathbf{A}(\mathbf{n},\mathbf{n}'')\\&=\mathbf{h}\cdot\mathbf{A}(-\mathbf{n}'',-\mathbf{n})\\&=\frac{ik}{2\pi}\left[\mathbf{h}\cdot\int_{S_a}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho})\exp(-ik\mathbf{n}\cdot\boldsymbol{\varrho})dS\right]\\&=\frac{ik}{2\pi}\left[\mathbf{h}''\cdot\int_{S_a}\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}}(\boldsymbol{\varrho})\exp(ik\mathbf{n}''\cdot\boldsymbol{\varrho})dS\right]\\&=-\frac{k^2}{2\pi}\int\int_{S_a}[\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho})]\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\cdot[\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}}(\boldsymbol{\varrho}')]dSdS', \quad (4\text{-}4)\end{aligned}$$

where $\mathbf{E}_{-\mathbf{n}}(\boldsymbol{\varrho})$ is the aperture field produced by an incident wave with propagation vector $-\mathbf{n}$ and magnetic polarization vector $\mathbf{h}''$. A suitable combination of these forms leads directly to a homogeneous expression for the scattered amplitude which is stationary with respect to independent variations of $\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}$ and $\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}}$ about the correct fields[11]

$$\frac{1}{\mathbf{h}''\cdot\mathbf{A}(\mathbf{n},\mathbf{n}'')}=\frac{1}{\mathbf{h}\cdot\mathbf{A}(-\mathbf{n}'',-\mathbf{n})}=\frac{-4\pi\int_{S_a}[\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho})]\cdot\Gamma^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\cdot[\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}}(\boldsymbol{\varrho}')]dSdS'}{\left[\mathbf{h}''\cdot\int_{S_a}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho})\exp(-ik\mathbf{n}\cdot\boldsymbol{\varrho})dS\right]\left[\mathbf{h}\cdot\int_{S_a}\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}}(\boldsymbol{\varrho})\exp(ik\mathbf{n}''\cdot\boldsymbol{\varrho})dS\right]}. \quad (4\text{-}5)$$

Accordingly, a homogeneous expression for t is

$$t(\mathbf{n}'')=\frac{-1}{2kS_a}Im\frac{\left[\mathbf{h}\cdot\int_{S_a}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho})\exp(-ik\mathbf{n}''\cdot\boldsymbol{\varrho})dS\right]\left[\mathbf{h}\cdot\int_{S_a}\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}''}(\boldsymbol{\varrho})\exp(ik\mathbf{n}''\cdot\boldsymbol{\varrho})dS\right]}{\int\int_{S_a}[\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho})]\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\cdot[\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}''}(\boldsymbol{\varrho}')]dSdS'}. \quad (4\text{-}6)$$

5. THE CIRCULAR APERTURE

If a plane wave is normally incident on a circular aperture of radius a, then it is evident that $S_a=\pi a^2$, $\mathbf{n}''=\mathbf{e}_z$, and $\mathbf{n}''\cdot\boldsymbol{\varrho}=0$. The symmetry of the circular aperture permits an arbitrary orientation of the incident magnetic polarization vector in the plane $z=0$, and it is convenient to choose $\mathbf{h}=\mathbf{e}_y$. Furthermore, in polar coordinates the circular aperture occupies the region

$$0\leq\rho\leq a;\quad 0\leq\varphi\leq 2\pi;\quad z=0.$$

With these specifications the stationary expression Eq. (4-6) for the transmission coefficient reduces to

$$t=-\frac{1}{2\pi ka^2}Im\frac{\left[\mathbf{e}_y\cdot\int_0^a\int_0^{2\pi}\rho d\rho d\varphi[\mathbf{e}_z\times E_{e_z}(\boldsymbol{\varrho})]\right]\left[\mathbf{e}_y\cdot\int_0^a\int_0^{2\pi}\rho d\rho d\varphi[\mathbf{e}_z\times\mathbf{E}_{-e_z}(\boldsymbol{\varrho})]\right]}{\int\int_0^a\int\int_0^{2\pi}\rho d\rho d\varphi\rho' d\rho' d\varphi'[\mathbf{e}_z\times\mathbf{E}_{e_z}(\boldsymbol{\varrho})]\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\cdot[\mathbf{e}_z\times\mathbf{E}_{-e_z}(\boldsymbol{\varrho}')]}. \tag{5-1}$$

Before choosing appropriate trial functions for the electric fields in the aperture, it will be worth while to examine some general properties of the aperture fields. The fields in question are produced by a pair of plane waves propagating in the directions $\mathbf{e}_z$ and $-\mathbf{e}_z$ with magnetic polarization vectors in the same direction parallel to the y axis. The tangetial component of the aperture field excited by the $\mathbf{e}_z$-directed wave can be assumed to be separable in polar coordinates, and to be given by

$$\mathbf{e}_z\times\mathbf{E}_{e_z}(\rho,\phi)=\mathbf{e}_z\times[\mathbf{e}_\rho R_\rho(\rho)\Phi_\rho(\phi)+\mathbf{e}_\phi R_\phi(\rho)\Phi_\phi(\phi)]. \tag{5-2}$$

Then, as a consequence of the symmetry of the incident waves and of the aperture, sketched in Fig. 2, it can be seen that the two tangential components of the aperture field associated with the $\mathbf{e}_z$-directed wave are

$$\begin{aligned}\mathbf{e}_\rho\cdot\mathbf{E}_{-e_z}(\rho,\pi-\phi)&=\mathbf{e}_\rho\cdot\mathbf{E}_{e_z}(\rho,\phi)=R_\rho(\rho)\Phi_\rho(\phi)\\ \mathbf{e}_\phi\cdot\mathbf{E}_{-e_z}(\rho,\pi-\phi)&=-\mathbf{e}_\phi\cdot\mathbf{E}_{e_z}(\rho,\phi)=-R_\phi(\rho)\Phi_\phi(\phi).\end{aligned} \tag{5-3}$$

The second field can, therefore, be expressed in terms of the components of the first, i.e.,

$$\mathbf{e}_z\times\mathbf{E}_{-e_z}(\rho,\pi-\phi)=\mathbf{e}_z\times[\mathbf{e}_\rho R_\rho(\rho)\Phi_\rho(\phi)-\mathbf{e}_\phi R_\phi(\rho)\Phi_\phi(\phi)]. \tag{5-4}$$

If it is assumed that the trial fields have the same ϕ-dependence as the incident field, then

$$\Phi_\rho(\phi)=\cos\phi,\quad \Phi_\phi(\phi)=-\sin\phi. \tag{5-5}$$

With Eqs. (2), (3), and (5) the aperture field functions at the point (ρ,ϕ) become

$$\begin{aligned}\mathbf{e}_z\times\mathbf{E}_{e_z}(\rho,\phi)&=\mathbf{e}_z\times[\mathbf{e}_\rho R_\rho(\rho)\cos\phi-\mathbf{e}_\phi R_\phi(\rho)\sin\phi]\\ \mathbf{e}_z\times\mathbf{E}_{-e_z}(\rho,\phi)&=\mathbf{e}_z\times[\mathbf{e}_\rho R_\rho(\rho)\cos(\pi-\phi)\\ &\qquad+\mathbf{e}_\phi R_\phi(\rho)\sin(\pi-\phi)],\end{aligned} \tag{5-6}$$

so that

$$\mathbf{e}_z\times\mathbf{E}_{e_z}(\rho,\phi)=-\mathbf{e}_z\times\mathbf{E}_{-e_z}(\rho,\phi)=\mathbf{e}_z\times\mathbf{E}(\rho,\phi).$$

A reasonable choice for the ρ dependence of the trial field is one which satisfies the same boundary conditions as those satisfied by the correct field. These conditions have been fully discussed by Meixner[2] and Bouwkamp[17] who show that at the rim of an aperture the tangential component of the electric field vanishes as $R^{\frac{1}{2}}$ and the normal component increases as $R^{-\frac{1}{2}}$, where R measures the distance from the field point to the edge. Accordingly, possible series representations of the radial functions associated with the aperture field components are

$$\begin{aligned}R_\rho(\rho)&=\frac{1}{\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}}\sum_{n=0}^{\infty}a_n\left(\frac{\rho}{a}\right)^{2n}\\ R_\phi(\rho)&=\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}\sum_{n=0}^{\infty}b_n\left(\frac{\rho}{a}\right)^{2n},\end{aligned} \tag{5-7}$$

where a_n, b_n are undetermined coefficients.

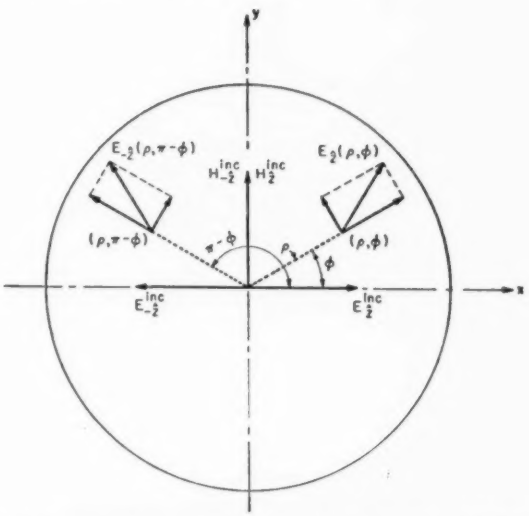


FIG. 2. Tangential aperture electric fields of opposite incidence with parallel magnetic polarization vectors.

[17] C. J. Bouwkamp, Physica **12**, 467 (1946).

With Eqs. (5) and (7), the ρ and ϕ components of the trial functions for the aperture field are

$$\mathbf{e}_\rho\cdot\mathbf{E}(\rho,\phi)=\frac{1}{\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}}\sum_{n=0}^{\infty}a_n\left(\frac{\rho}{a}\right)^{2n}\cos\phi$$

$$\mathbf{e}_\phi\cdot\mathbf{E}(\rho,\phi)=\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}\sum_{n=0}^{\infty}b_n\left(\frac{\rho}{a}\right)^{2n}\sin\phi. \tag{5-8}$$

The corresponding rectangular components are

$$\left.\begin{aligned}\mathbf{e}_x\cdot\mathbf{E}(\rho,\phi)=&\frac{1}{\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}}\sum_{n=0}^{\infty}a_n\left(\frac{\rho}{a}\right)^{2n}\cos^2\phi\\&-\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}\sum_{n=0}^{\infty}b_n\left(\frac{\rho}{a}\right)^{2n}\sin^2\phi\\\mathbf{e}_y\cdot\mathbf{E}(\rho,\phi)=&\left[\frac{1}{\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}}\sum_{n=0}^{\infty}a_n\left(\frac{\rho}{a}\right)^{2n}\right.\\&\left.\left(-1\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}\sum_{n=0}^{\infty}b_n\left(\frac{\rho}{a}\right)^{2n}\right]\\&\times\sin\phi\cos\phi\end{aligned}\right\}. \tag{5-9}$$

At the center of the aperture these components reduce to

$$\mathbf{e}_x\cdot\mathbf{E}(\rho,\phi)=(a_0+b_0)\cos^2\phi-b_0$$

$$\mathbf{e}_y\cdot\mathbf{E}(\rho,\phi)=(a_0+b_0)\sin\phi\cos\phi.$$

Since the field is a single-valued function of the coordinates, it should be independent of ϕ when $\rho=0$. It is necessary, therefore, to require that

$$a_0=-b_0.$$

This condition also insures that $\nabla\cdot\mathbf{E}=0$ at the center of the aperture.

It may be noted here that if the series are truncated so that $a_n=0$ for $n\geqslant 2$, $b_n=0$ for $n\geqslant 1$, and the remaining coefficients are assigned the values $a_1=-a$, $a_0=-b_0=2a$, the trial function reduces to

$$\mathbf{e}_\rho\cdot\mathbf{E}(\rho,\phi)=\frac{2a^2-\rho^2}{(a^2-\rho^2)^{\frac{1}{2}}}\cos\phi.$$

$$\mathbf{e}_\phi\cdot\mathbf{E}(\rho,\phi)=-2(a^2-\rho^2)^{\frac{1}{2}}\sin\phi.$$

These functions constitute the low-frequency exact solution obtained by Bouwkamp.[4] They are frequency independent and are valid only for $ka\ll 1$. In reference 11, they are used as a zeroth-order trial function by Levine and Schwinger.

The vector trial function $\mathbf{e}_z\times\mathbf{E}(\varrho)$ of Eq. (5-8) contains undetermined frequency-dependent coefficients. These may be chosen appropriately by using the stationary property of the expression for the transmission coefficient. In accordance with Eqs. (4-2) and (5-1) the stationary form of interest is

$$t=\frac{2}{ka^2}ImA_y=-\frac{2}{ka^2}Im\frac{I_1^2}{4\pi I_2}. \tag{5-10}$$

Substitution of the series for $\mathbf{e}_z\times\mathbf{E}(\varrho)$ shows at once that

$$I_1=\int_0^a\int_0^{2\pi}\rho d\rho d\phi\,\mathbf{e}_y\cdot\mathbf{e}_z\times\mathbf{E}(\varrho)=\sum_{n=0}^{\infty}(a_nB_n{}^a+b_nB_n{}^b)$$

$$I_2=\int_0^a\int_0^{2\pi}\rho d\rho d\phi\rho' d\rho' d\phi'\,\mathbf{e}_z\times\mathbf{E}(\varrho)\cdot\mathbf{\Gamma}^{(0)}(\rho,\phi;\rho',\phi')\cdot\mathbf{e}_z\times\mathbf{E}(\varrho')$$
$$=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}[a_na_mC_{nm}{}^{aa}+a_nb_mC_{nm}{}^{ab}+b_nb_mC_{nm}{}^{bb}],$$

where

$$B_n{}^a=\pi\int_0^a\frac{\rho d\rho}{\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}}\left(\frac{\rho}{a}\right)^{2n}$$

$$B_n{}^b=-\pi\int_0^a\rho\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}\left(\frac{\rho}{a}\right)^{2n}d\rho$$

$$C_{nm}{}^{aa}=\int_{S_1}dSdS'\frac{\left(\frac{\rho}{a}\right)^{2n}\left(\frac{\rho'}{a}\right)^{2m}}{\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}\left(1-\left(\frac{\rho'}{a}\right)^2\right)^{\frac{1}{2}}}\times\cos\phi\cos\phi'\Gamma_{\phi\phi'}{}^{(0)}(\rho,\phi;\rho',\phi')=C_{mn}{}^{aa}$$

$$C_{nm}{}^{ab}=-\int_{S_1}dSdS'\frac{\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}}{\left(1-\left(\frac{\rho'}{a}\right)^2\right)^{\frac{1}{2}}}\left(\frac{\rho}{a}\right)^{2n}\left(\frac{\rho'}{a}\right)^{2m}\times\sin\phi\cos\phi'\Gamma_{\rho\phi'}{}^{(0)}(\rho,\phi;\rho',\phi')$$
$$-\int_{S_1}dSdS'\frac{\left(1-\left(\frac{\rho'}{a}\right)^2\right)^{\frac{1}{2}}}{\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}}\left(\frac{\rho}{a}\right)^{2n}\left(\frac{\rho'}{a}\right)^{2m}\times\cos\phi\sin\phi'\Gamma_{\phi\rho'}{}^{(0)}(\rho,\phi;\rho',\phi')=C_{mn}{}^{ab}$$

$$C_{nm}{}^{bb}=\int_{S_1} dSdS'\left(1-\left(\frac{\rho}{a}\right)^2\right)^{\frac{1}{2}}\left(1-\left(\frac{\rho'}{a}\right)^2\right)^{\frac{1}{2}}$$
$$\times\left(\frac{\rho}{a}\right)^{2n}\left(\frac{\rho'}{a}\right)^{2m}\sin\phi\,\sin\phi'\Gamma_{\rho\rho'}{}^{(0)}(\rho,\phi;\rho',\phi')$$
$$=C_{mn}{}^{bb}.$$

The $\Gamma_{uv}{}^{(0)}$ are the $\mathbf{e}_u\mathbf{e}_v$ components of the free-space dyadic Green's function. With these substitutions the scattered amplitude A_y takes the form

$$A_y=\frac{-1}{4\pi}\frac{\sum_{n=0}^{\infty}(a_nB_n{}^a+b_nB_n{}^b)^2}{\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}[a_na_mC_{nm}{}^{aa}+a_nb_mC_{nm}{}^{ab}+b_nb_mC_{nm}{}^{bb}]},$$

where the B's and C's are, in principle, known functions of the wavelength and the size of the aperture, while the a's and b's remain to be determined.

The procedure is to differentiate A_y with respect to each of the independent coefficients and set

$$\frac{\partial A_y}{\partial a_\nu}=\frac{\partial A_y}{\partial b_\mu}=0.$$

Since the behavior of the trial function at the center of the aperture requires $a_0=-b_0$, extremizing A_y with respect to either of these coefficients gives

$$A_y\sum_{m=0}^{\infty}[2a_mC_{0m}{}^{aa}-(a_m-b_m)C_{0m}-2b_mC_{0m}{}^{bb}]$$
$$=\frac{-1}{2\pi}(B_0{}^a-B_0{}^b)\sum_{m=0}^{\infty}[a_mB_m{}^a+b_mB_m{}^b].$$

But by definition

$$A_y=\frac{ik}{2\pi}I_1=\frac{ik}{2\pi}\sum_{m=0}^{\infty}(a_mB_m{}^a+b_mB_m{}^b), \qquad (5\text{-}11)$$

so that

$$\sum_{m=0}^{\infty}[(2C_{0m}{}^{aa}-C_{0m}{}^{ab})a_m-(2C_{0m}{}^{bb}-C_{0m}{}^{ab})b_m]$$
$$=\frac{i}{k}(B_0{}^a-B_0{}^b). \qquad (5\text{-}12)$$

A similar treatment with respect to a_ν and b_μ gives

$$\sum_{m=0}^{\infty}[2C_{\nu m}{}^{aa}a_m+C_{\nu m}{}^{ab}b_m]=\frac{i}{k}B_\nu{}^a \quad \nu=1, 2, \cdots \qquad (5\text{-}13)$$

$$\sum_{m=0}^{\infty}[C_{\mu m}{}^{ab}a_m+2C_{\mu m}{}^{bb}b_m]=\frac{i}{k}B_\mu{}^b \quad \mu=1, 2, \cdots. \qquad (5\text{-}14)$$

Equations (12), (13), and (14) constitute the linear set of algebraic equations required to determine the a_m and b_m. With these, the transmission coefficient can be calculated through Eq. (11).

In a practical calculation the trial function must be terminated at some convenient value of m. The approximate transmission coefficient obtained in this way will be

$$t^{(N)}=-\frac{1}{\pi a^2}\,\mathrm{Re}[\sum_{m=0}^{j}a_mB_m{}^a+\sum_{m=0}^{k}b_mB_m{}^b], \qquad (5\text{-}15)$$

where $N=j+k$ is the order of the approximation. For a first-order solution with $j=1$, $k=0$, the set of equations to be solved consists only of Eq. (12), and Eq. (13) with $\nu=1$:

$$2(C_{00}{}^{aa}-C_{00}{}^{aa}+C_{00}{}^{bb})a_0+(2C_{01}{}^{aa}-C_{01}{}^{ab})a_1$$
$$=\frac{i}{k}(B_0{}^a-B_0{}^b) \qquad (5\text{-}16)$$

$$(2C_{01}{}^{aa}-C_{01}{}^{ab})a_0+2C_{11}{}^{aa}a_1=\frac{i}{k}B_1{}^a. \qquad (5\text{-}17)$$

The coefficients a_0 and a_1 are readily determined from these equations, and when they are put into Eq. (15) there results

$$t^{(1)}(\alpha)=\frac{32\alpha}{9\pi}Im\frac{4P(\alpha)-2Q(\alpha)+R(\alpha)}{Q^2(\alpha)-4P(\alpha)R(\alpha)}. \qquad (5\text{-}18)$$

Here, $\alpha=ka$ and

$$P(\alpha)=C_{11}{}^{aa}$$
$$Q(\alpha)=2C_{01}{}^{aa}-C_{01}{}^{ab}$$
$$R(\alpha)=C_{00}{}^{aa}-C_{00}{}^{ab}+C_{00}{}^{bb}.$$

The numerical factors in formula Eq. (18) are the result of evaluating the B's. These are

$$B_0{}^a=\pi a\int_0^a\frac{\rho d\rho}{(a^2-\rho^2)^{\frac{1}{2}}}=\pi a^2$$

$$B_1{}^a=\frac{\pi}{a}\int_0^a\frac{\rho^3d\rho}{(a^2-\rho^2)^{\frac{1}{2}}}=\tfrac{2}{3}\pi a^2$$

$$B_0{}^b=-\frac{\pi}{a}\int_0^a\rho(a^2-\rho^2)^{\frac{1}{2}}d\rho=-\tfrac{1}{3}\pi a^2.$$

The final expressions from which P, Q, and R may be calculated are given in the appendix.

The computed values of $t^{(1)}$ are compared with previous results in Fig. 3. It can be seen that the variational approximations are quite close to the exact solution[18] and that $t^{(1)}$ provides a much higher degree of accuracy than $t^{(0)}$ in the range $1.5\leqslant ka\leqslant 4.5$. A further comparison of interest can be made for small values

[18] W. Andrejewski, Z. angew. Physik **5**, 178 (1953).

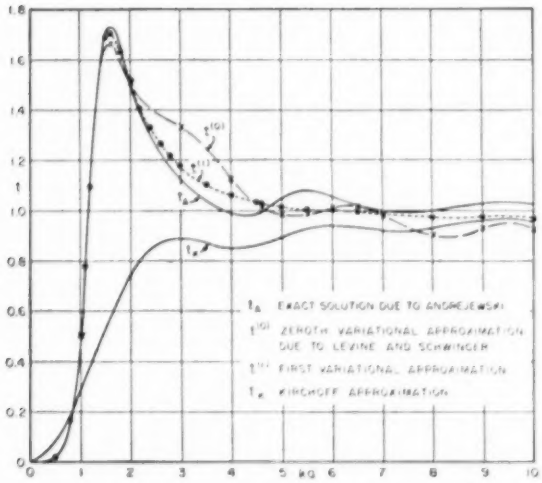

t_A EXACT SOLUTION DUE TO ANDREJEWSKI
$t^{(0)}$ ZEROTH VARIATIONAL APPROXIMATION DUE TO LEVINE AND SCHWINGER
$t^{(1)}$ FIRST VARIATIONAL APPROXIMATION
t_K KIRCHOFF APPROXIMATION
ka

FIG. 3. Transmission coefficient of a circular aperture for normal incidence.

of ka. If the transmission coefficients obtained by the various theories are expanded into power series in ka, the results are:

(i) Zeroth-order approximation by the variational method (due to Levine and Schwinger[11])

$$t^{(0)}=\frac{64(ka)^4}{27\pi^2}\left[1+\frac{22}{25}(ka)^2+0.4079(ka)^4+\cdots\right]. \quad (5\text{-}19)$$

(ii) First-order approximation by the variational method (obtained in this report)

$$t^{(1)}=\frac{64(ka)^4}{27\pi^2}\left[1+\frac{22}{25}(ka)^2+0.3968(ka)^4+\cdots\right]. \quad (5\text{-}20)$$

(iii) Exact solution (due to Bouwkamp)[4]

$$t_B=\frac{64(ka)^4}{27\pi^2}\left[1+\frac{22}{25}(ka)^2+0.3979(ka)^4+\cdots\right]. \quad (5\text{-}21)$$

6. THE ELLIPTICAL APERTURE

If the aperture in the plane screen is elliptical in shape, the general procedure of section 5 is much more difficult to apply. Either elliptic or oblate spheroidal coordinates must be used, and the computational labor would be greatly increased. However, considerable improvement in the Kirchhoff result can be effected by using a single-component trial function for the electric field in the aperture corresponding to the original trial function of Levine and Schwinger in the circular case.

For a normally incident plane wave polarized so that $\mathbf{E}^{\text{inc}}(\mathbf{r})=\mathbf{e}_x e^{ikz}$ and $\mathbf{H}^{\text{inc}}(\mathbf{r})=\mathbf{e}_y e^{ikz}$, the choice of trial function is‡

$$\mathbf{E}(\rho)=\mathbf{e}_x(1-\rho^2)^{\frac{1}{2}}. \quad (6\text{-}1)$$

‡ The elliptical polar coordinates $\rho=\left(\left(\frac{x}{a}\right)^2+\left(\frac{y}{b}\right)^2\right)^{\frac{1}{2}}$, $\phi=\tan^{-1}\frac{ay}{bx}$ are used here.

Substitution of this function into the stationary expression for the transmission coefficient gives

$$t=\frac{-2\pi ab}{9k}Im\frac{1}{I}, \quad (6\text{-}2)$$

where

$$I=\int_0^1\int_0^{2\pi}\rho d\rho d\phi\rho' d\rho' d\phi' \times(1-\rho^2)^{\frac{1}{2}}(1-\rho'^2)^{\frac{1}{2}}\Gamma_{yy}^{(0)}(\rho,\phi;\rho',\phi')$$

With the representation

$$\Gamma_{yy}^{(0)}(\rho,\phi;\rho',\phi')=\frac{1}{8\pi}\int_{-\infty}^{\infty}\frac{dk_x dk_y}{(k^2-k_x^2-k_y^2)^{\frac{1}{2}}}\left(1-\frac{k_y^2}{k^2}\right) \times\exp[ik_x(x-x')+ik_y(y-y')],$$

the integral can be rearranged, following reference (16), so that it takes the form

$$I=\frac{a^2b^2\pi}{4k^2}\left\{\int_0^{2\pi}\int_0^{\infty}\frac{dvd\theta(v^2-1)^{\frac{1}{2}}J_1^2(kLv)}{L^3v^2} -\int_0^{2\pi}\int_0^{\infty}dvd\theta\frac{\cos^2\theta J_1^2(kLv)}{L^3(v^2-1)^{\frac{1}{2}}}\right\}, \quad (6\text{-}3)$$

where

$$L=[(a\cos\theta)^2+(b\sin\theta)^2]^{\frac{1}{2}}.$$

Finally, by carrying out the integration with respect to v, we find that§

$$I=\frac{-a^2b^2\pi}{4k^2}\int_0^{2\pi}[F(kL)-\cos^2\theta G(kL)]\frac{d\theta}{L^3}, \quad (6\text{-}4)$$

where

$$F(x)=\frac{1}{4x}\left\{\left[\frac{-1}{2x}J_0(2x)-J_1(2x)+\left(1+\frac{1}{4x^2}\right)\int_0^{2x}J_0(t)dt\right] -i\left[\frac{2x^2}{\pi}-\frac{1}{\pi}+\frac{1}{2x}S_0(2x)-\left(1+\frac{1}{4x^2}\right)\int_0^{2x}S_0(t)dt\right]\right\}$$

$$G(x)=\frac{1}{x}\left\{\left[\frac{-1}{2x}J_0(2x)-J_1(2x)+\left(\frac{1}{2}+\frac{1}{4x^2}\right)\int_0^{2x}J_0(t)dt\right. +i\left.\left[\frac{1}{\pi}-\frac{1}{2x}S_0(2x)-S_1(2x)+\left(\frac{1}{2}+\frac{1}{4x^2}\right)\int_0^{2x}S_0(t)dt\right]\right\}.$$

Numerical values of the transmission coefficient are given in Fig. 4 for $0<ka\leqslant 10$, and $b/a=\frac{1}{2}, \frac{1}{3}$. These are plotted together with the corresponding result for the circular aperture which has been given in reference (11), and which can be obtained from Eq. (2) by

§ The first integral is related to $F_{11}(kL)$ in reference 19; the second is related to $I_2(kL)$ in appendix A of reference 20.

putting $b/a=1$. Using the expansions

$$F(x)=\left[\frac{1}{3}-\frac{x^2}{15}+\frac{x^4}{140}+\cdots\right]-i\left[\frac{2x^3}{27\pi}-\frac{4x^5}{675\pi}+\frac{16x^7}{55\,125\pi}+\cdots\right] \quad (6\text{-}5)$$

$$G(x)=\left[\frac{1}{3}+\frac{x^2}{15}-\frac{3x^4}{140}+\cdots\right]+i\left[\frac{4x^3}{27\pi}-\frac{16x^5}{675\pi}+\frac{32x^7}{18\,375\pi}+\cdots\right]$$

in Eq. (4), a form of the transmission coefficient appropriate to small values of ka can be obtained:

$$t=\frac{64}{27\pi^2}\left(\frac{a}{b}\right)(ka)^4\left[A_0\left(\frac{b}{a}\right)+A_2\left(\frac{b}{a}\right)(ka)^2+A_4\left(\frac{b}{a}\right)(ka)^4+\cdots\right], \quad (6\text{-}6)$$

where $A_0(b/a)$, $A_2(b/a)$, $A_4(b/a)$, $\cdots$ are functions of the eccentricity, $e=(1-(b/a)^2)^{\frac{1}{2}}$. They are given in appendix B of reference 20. With small eccentricity Eq. (6) becomes

$$t=\frac{64}{27\pi^2}\left(\frac{a}{b}\right)(ka)^4\left[\left(1-\frac{9}{4}e^2\cdots\right)+\left(\frac{27}{25}-\frac{349}{100}e^2+\cdots\right)(ka)^2+\left(\frac{8937}{12\,250}-\frac{155\,223}{49\,000}e^2+\cdots\right)(ka)^4+\cdots\right]. \quad (6\text{-}7)$$

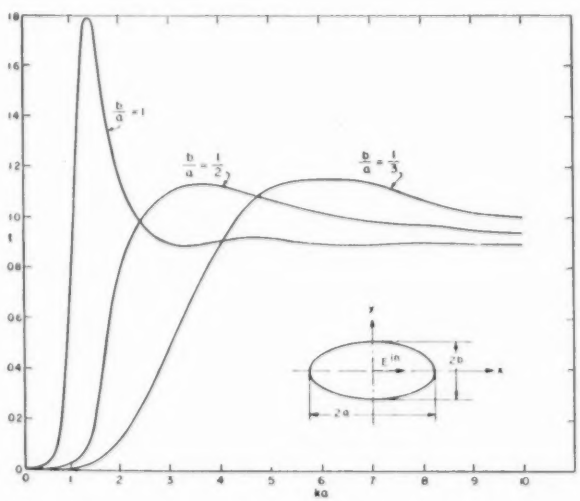


FIG. 4. Transmission coefficient of elliptical apertures for normal incidence, a variational approximation based on a tangential electric aperture field of $\mathbf{e}_z\left[1-\left(\frac{x}{a}\right)^2-\left(\frac{y}{b}\right)^2\right]^{\frac{1}{2}}$.

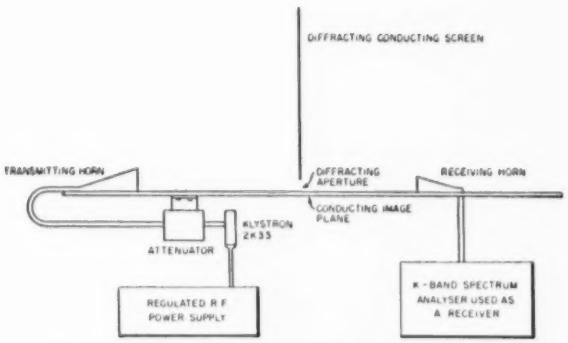


FIG. 5. Microwave components for measuring aperture transmission coefficients.

7. THE MEASUREMENT OF APERTURE TRANSMISSION COEFFICIENTS

Description of the Apparatus

A schematic diagram of the experimental arrangement used to measure transmission coefficients is shown in Fig. 5. In its original form the apparatus was designed and built for the investigation of diffraction by cylinders under conditions that approached the idealizations of theory as closely as possible.[15] The general features of the present problem that must be approximated physically are:

1. That the perforated screen is isolated in space.
2. That the incident wave is plane.
3. That the perforated screen has zero thickness.

In this experiment the isolation of the screen is approximated by mounting it vertically over a conducting image plane. The image plane, which is 96 wavelengths wide by 144 wavelengths long at $\lambda=1.25$ cm, is large enough so that reflections from its edges are negligible. To make the incident wave nearly plane, at least over the region of the aperture, a remote point source is required. In this apparatus such a source is approximated adequately with a horn radiator. The horn is driven by a 2K33 klystron through a length of K-band wave guide and provides the power gain required for measuring transmission coefficients. The thin screen is made of a chrome-plated steel sheet about 0.02 wavelength thick. Its lower edge is fastened securely to the image plane; the upper edge is attached to a support, shown in Fig. 6, by means of which tension can be applied uniformly along the edge to keep the thin screen vertical and plane. The center section of the screen contains the half-aperture and is removable so that apertures of different sizes and shapes can be put in place with a minimum of effort.

With less than 20 milliwatts available from a 2K33 K-band oscillator very little energy is transmitted by a small aperture located far from the transmitting horn. Consequently, the measurement of its transmission coefficient, which is related to the far-zone field on the shadow side of the screen, requires a very sensitive

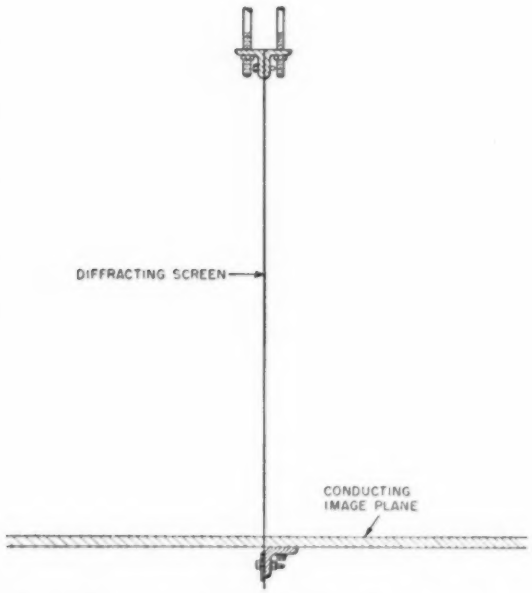


FIG. 6. Supporting structure for the diffracting screen.

detector and corresponding care in preventing leakage to the detector by paths other than through the aperture. A simple criterion is that with the aperture closed the leakage signal must be less than noise; the requirement is met as follows: 1. The screen containing the aperture is made large enough so that no detectable radiation is diffracted around any of its three free edges. 2. The fourth (bottom) edge of the screen is pinched tightly in a narrow slit in the image plane so that the joint between the two planes is not leaky. 3. The aperture insert is put into intimate contact with the rest of the diffracting screen by covering the overlapping joint with strips of aluminum foil two wavelengths wide and 0.0012 wavelength thick. These strips are sealed with conducting silver paint (see Figs. 7 and 8).

Under these conditions no leakage signal can be detected by the receiving system, which consists of an antenna, a calibrated attenuator, and a spectrum-analyzer. The receiving antenna is a horn located 30 wavelengths from the diffracting aperture in the direction of incidence. It is designed for optimum directivity and flares 19 degrees in the E plane and 22.5 degrees in the H plane from standard K-band wave guide to a rectangular aperture 11 wavelengths wide and three wavelengths high. From the horn the signal is transmitted by wave guide to a K-band spectrum-analyzer (TSK 2SE) through a precision variable attenuator. The combination of attenuator and spectrum-analyzer is used as a sensitive receiver.

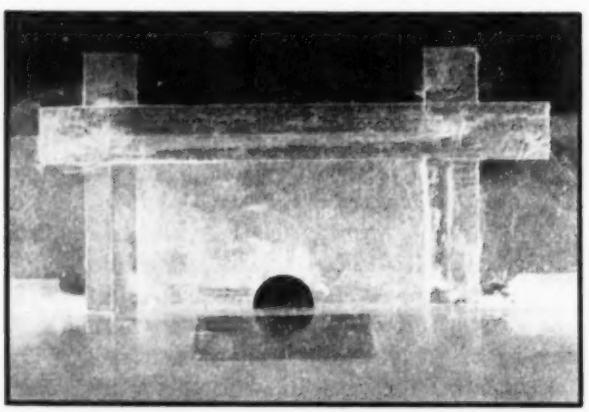


FIG. 7. Circular aperture mounted over image plane.

Method of Measurement

The analytic definition of the transmission coefficient of an aperture is

$$t=\frac{\mathrm{Re}\int_{S_a}\mathbf{e}_z\cdot\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\rho}')\times\mathbf{H}_{\mathbf{n}''}{}^{*}(\boldsymbol{\rho}')dS'}{S_a|\mathbf{E}^{\mathrm{inc}}(\mathbf{r})\times\mathbf{H}^{\mathrm{inc}*}(\mathbf{r})|}, \tag{7-1}$$

where S_a is the area of the aperture and the subscript $\mathbf{n}''$ denotes the direction of propagation of the incident plane wave,

$$\mathbf{E}^{\mathrm{inc}}(\mathbf{r})=\mathbf{e}E_0\exp(ik\mathbf{n}''\cdot\mathbf{r}),\quad \mathbf{H}^{\mathrm{inc}}(\mathbf{r})=\mathbf{h}H_0\exp(ik\mathbf{n}''\cdot\mathbf{r}).$$

Since the aperture field distribution cannot be measured easily, this definition is not very good in the operational sense. A better one is made available by the analysis given in section 4, which makes it possible to put Eq. (1) into a form involving only a far-zone electric field. First, however, by changing the order of the vector double product in the integrand of Eq. (1), it is possible to make use of the fact that in the aperture $H_{\tan}(\boldsymbol{\rho})=H_{\tan}{}^{\mathrm{inc}}(\boldsymbol{\rho})$. The formula then simplifies to

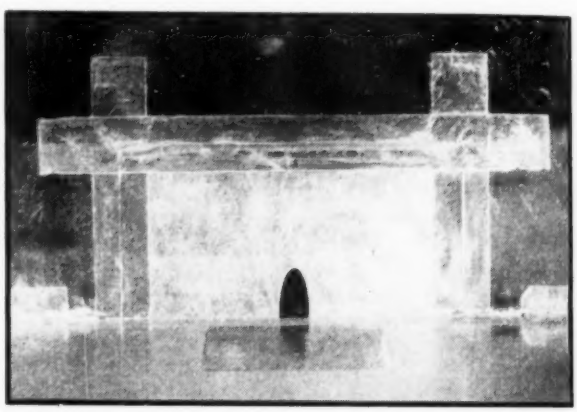


FIG. 8. Elliptical aperture mounted over image plane.

$$t=\frac{\mathrm{Re}\int_{S_a}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\rho}')\cdot\mathbf{h}\exp(-ik\mathbf{n}''\cdot\boldsymbol{\rho}')dS'}{S_aE_0}. \tag{7-2}$$

The integral in Eq. (2) can be interpreted readily in terms of the far-zone electric field on the shadow side of the screen. The asymptotic form of this field is

$$\mathbf{E}_{\mathbf{n}''}(\mathbf{r})=\mathbf{n}\times\mathbf{A}(\mathbf{n},\mathbf{n}'')\frac{e^{ikr}}{r},\quad r\to\infty, \tag{7-3}$$

where

$$\mathbf{A}(\mathbf{n},\mathbf{n}'')=\frac{ik}{2\pi}\int_{S_a}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho}')\exp(-ik\mathbf{n}\cdot\boldsymbol{\varrho}')dS'.$$

Comparison of this expression with Eq. (2) shows that in terms of the scattered amplitude in the direction of incidence the transmission coefficient takes the simple form,

$$t=\frac{2\pi}{kS_aE_0}Im[\mathbf{h}\cdot\mathbf{A}(\mathbf{n}'',\mathbf{n}'')]. \tag{7-4}$$

With normal incidence $\mathbf{n}''=\mathbf{e}_z$, $\mathbf{h}=\mathbf{e}_y$, and Eq. (4)

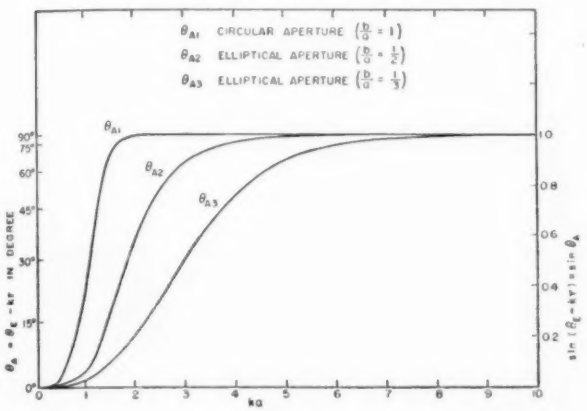


FIG. 9. Phase angle of the far-zone scattered amplitude computed by variational methods.

becomes

$$t=\frac{2\pi}{kS_aE_0}Im[(\mathbf{e}_y\cdot\mathbf{A}(\mathbf{e}_z,\mathbf{e}_z)]. \tag{7-5}$$

This formula can be written in terms of the scattered electric field by setting $\mathbf{r}=\mathbf{e}_zr$ in Eq. (3) and forming the scalar product with the unit vector $\mathbf{e}_x$. The result is

$$\mathbf{e}_y\cdot\mathbf{A}(\mathbf{e}_z,\mathbf{e}_z)=\frac{r\mathbf{e}_x\cdot\mathbf{E}(\mathbf{e}_zr)}{e^{ikr}}. \tag{7-6}$$

Accordingly,

$$t=\frac{2\pi r}{kS_aE_0}|E_x(\mathbf{e}_yr)|\sin(\theta_E-kr), \tag{7-7}$$

where θ_E is the phase angle of the complex quantity $E_x(\mathbf{e}_zr)$.

This form of the transmission coefficient makes it evident that it can be determined experimentally from a measurement of the amplitude and phase of E_x in the far zone of the aperture. Unfortunately, with the small amount of power available the phase measurement is not feasible, but reasonable values of t can still be obtained by measuring $|E_x|$ and estimating θ_E from the theoretical results of the previous sections. It is found that over most of the frequency range of interest $\sin(\theta_E-kr)$ is nearly unity as shown in Fig. 9. This value checks with the optical limit of the exact theory.

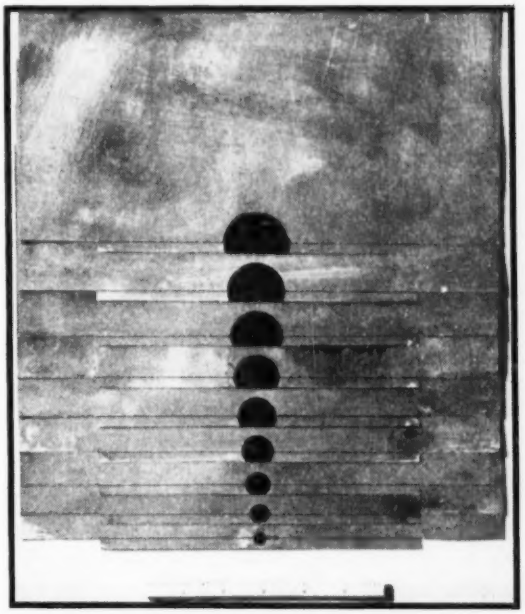


FIG. 10. Circular apertures.

With the definitions just given, the measurement of transmission coefficients for normal incidence is reduced

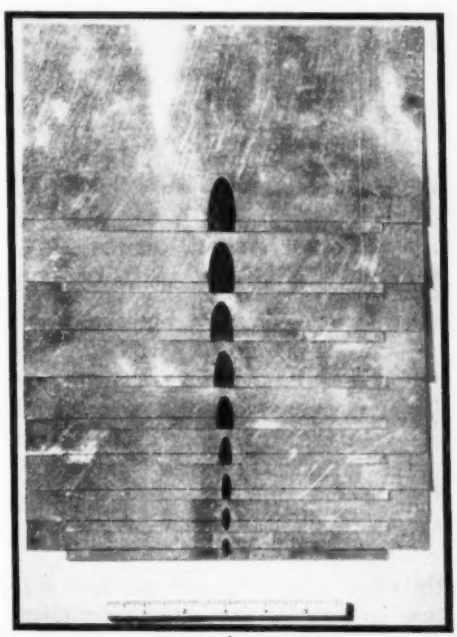


FIG. 11. Elliptical apertures $b/a=\frac{1}{3}$.

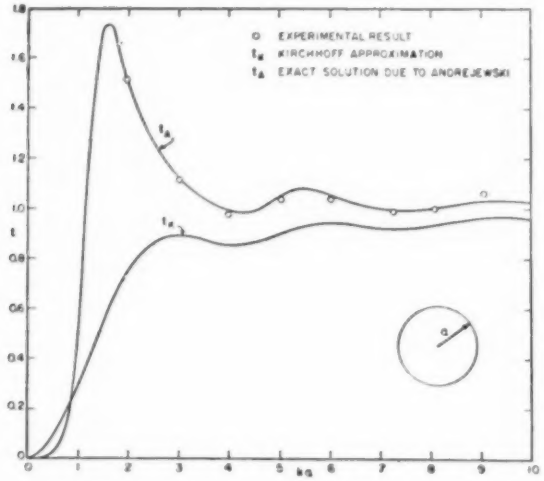


FIG. 12. Measured transmission coefficient of circular apertures for normal incidence.

to the problem of measuring the far-zone electric field directly behind the aperture. Since this field is small and frequency sensitive, the radiation frequency must be known and held constant throughout a series of measurements. Then the actual measurement of the amplitude of the far-zone electric field for each aperture is a relatively simple matter. The diffracted wave that is incident upon the receiving horn is proportional to $|E_x|$ at a point directly behind the aperture. The resulting signal is transmitted through a section of wave guide containing a precision attenuator to a spectrum analyzer where it is displayed on an oscilloscope. By adjusting the attenuator so that, with each aperture, the signal is constant in amplitude the relative magtude of E_x can be determined from the attenuator settings at each value of ka. The relative transmission coefficients of a group of similar apertures are then readily calculated from Eq. (7), and they can be compared with theoretical results after being normalized to the limiting value of unity for large values of ka.

Some of the apertures for which this measuring procedure was carried out are shown in Figs. 10 and 11.

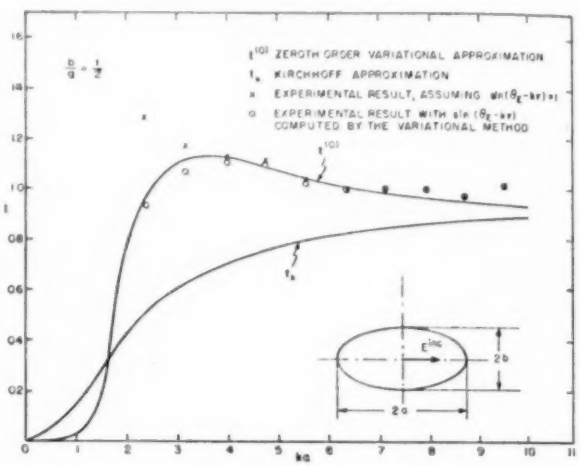
Comparison of Results

The experimental values of the transmission coefficient for circular apertures are compared to the results of the exact theory in Fig. 12. Good agreement is obtained over almost the entire frequency range considered. The largest error occurs at $ka=9$, and is probably due to the fact that over a large aperture the incident wave from a point source is no longer plane to the required degree of approximation. The transmitted signal was too small for reliable measurements to be made in the region $ka<2$.

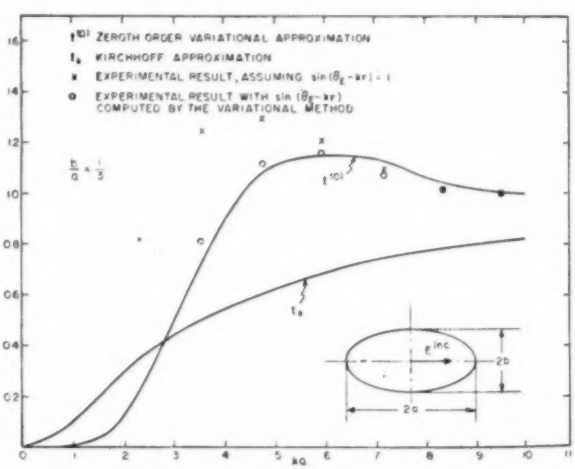
Since the exact theory has not yet been worked out for elliptical apertures, the experimental results are compared to those obtained by the variational and Kirchhoff approximations.[16] The curves are shown in Fig. 13. The remarkable agreement between the experimental result and the variational approximation probably indicates that the single-component trial field resembles the true field in the aperture quite closely when the incident electric field is polarized along the major axis of the ellipse. The contrast with the Kirchhoff approximation hardly needs to be emphasized; this approximation tends toward the correct result only for large values of ka.

8. CONCLUSION

In a recent review of the problem of diffraction through a circular aperture Andrejewski[18] states that he considers the usefulness of the variational method to be severely limited by the necessity of using extremely simple trial functions. He implies that no more computational labor is involved in obtaining the rigorous solution to a diffraction problem than in finding a variational approximation from a trial function more exact than the static distribution, supposing such a trial function can be found in the first place.



(a)



(b)

FIG. 13(a) and (b). Measured transmission coefficients of elliptical apertures for normal incidence.

One of the aims of this paper has been to show that in the vector, as well as the scalar diffraction problem, a systematic and well-defined procedure is available for improving the static distribution. This procedure leads to a result for the transmission coefficient of a circular aperture that reproduces the first maximum with a very high degree of accuracy. Moreover, for no frequency in the range $ka<10$ is the error more than 10 percent; usually it is much less than that. This gain over previous approximations is achieved without the additional computations becoming excessive since only well-tabulated functions are involved.

It is true that the ultimate value of the variational procedure cannot be judged on the basis of its ability to yield better and better approximations to results that have already been obtained by exact methods. But this criticism no longer applies when we deal with the elliptical aperture. Since no study has been made of ellipsoidal functions, an exact solution is not available for this shape and would probably be exceedingly difficult to carry out in any case. Yet the variational formulation, with the crudest trial function, readily provides results that agree well with experiment over the entire range of ka that is accessible to measurement. It is in this direction that the method can be exploited with advantage.

ACKNOWLEDGMENTS

The research reported in this paper was made possible through support extended to Cruft Laboratory, Harvard University, jointly by the Navy Department (Office of Naval Research), the Signal Corps of the U. S. Army, and the U. S. Air Force under ONR Contract N5ori-76.

The authors wish to thank Miss Lesley Gerould and Mrs. Mary Andrews who carried out the numerical computations.

APPENDIX

The function $P(\alpha)$, $Q(\alpha)$, and $R(\alpha)$ of Eq. (5-18) are evaluated in reference 16. In terms of Bessel functions of the first kind and Struve functions‖ they are

$$P(\alpha)=\frac{\pi^2a^3}{4\alpha^2}\left\{\left[\left(2-\frac{11}{8}\frac{1}{\alpha^2}\right)J_0(2\alpha)-\frac{11}{4}\frac{1}{\alpha}J_1(2\alpha)+\left(\frac{\alpha}{2}-\frac{1}{4}\frac{1}{\alpha}+\frac{11}{16}\frac{1}{\alpha^3}\right)\int_0^{2\alpha}J_0(t)dt\right]+i\left[\frac{-3\alpha}{2\pi}+\frac{11}{4\pi}\frac{1}{\alpha}+\left(2-\frac{11}{8}\frac{1}{\alpha^2}\right)S_0(2\alpha)-\frac{11}{4}\frac{1}{\alpha}S_1(2\alpha)+\left(\frac{\alpha}{2}-\frac{1}{4}\frac{1}{\alpha}+\frac{11}{16}\frac{1}{\alpha^3}\right)\int_0^{2\alpha}S_0(t)dt\right]\right\}$$

$$Q(\alpha)=\frac{\pi^2a^3}{4\alpha^2}\left\{\left[\left(3-\frac{1}{\alpha^2}\right)J_0(2\alpha)-\frac{2}{\alpha}J_1(2\alpha)+\left(\alpha-\frac{3}{2}\frac{1}{\alpha}+\frac{1}{2\alpha^3}\right)\int_0^{2\alpha}J_0(t)dt\right]+i\left[\frac{-2\alpha}{\pi}+\frac{2}{\pi\alpha}+\left(3-\frac{1}{\alpha^2}\right)S_0(2\alpha)-\frac{2}{\alpha}S_1(2\alpha)+\left(\alpha-\frac{3}{2}\frac{1}{\alpha}+\frac{1}{2\alpha^3}\right)\int_0^{2\alpha}S_0(t)dt\right]\right\}$$

$$R(\alpha)=\frac{\pi^2a^3}{4\alpha^2}\left\{\left[\left(1+\frac{1}{8}\frac{1}{\alpha^2}\right)J_0(2\alpha)+\frac{1}{4\alpha}J_1(2\alpha)+\left(\frac{\alpha}{2}-\frac{5}{4}\frac{1}{\alpha}-\frac{1}{16}\frac{1}{\alpha^3}\right)\int_0^{2\alpha}J_0(t)dt\right]+i\left[\frac{\alpha}{2\pi}-\frac{1}{4\pi\alpha}+\left(1+\frac{1}{8}\frac{1}{\alpha^2}\right)S_0(2\alpha)+\frac{1}{4\alpha}S_1(2\alpha)+\left(\frac{\alpha}{2}-\frac{5}{4}\frac{1}{\alpha}-\frac{1}{16}\frac{1}{\alpha^3}\right)\int_0^{2\alpha}S_0(t)dt\right]\right\}.$$

‖ A detailed discussion of the Struve function can be found in G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, Cambridge, 1945), p. 328. Table I of this volume includes a tabulation of $S_0(x)$ and $S_1(x)$.

Relation of Antimony Transmission and the Photoelectric Yield of Cs–Sb

MARTIN ROME
Westinghouse Electric Corporation, Elmira, New York
(Received April 5, 1954)

The relation of transmission of thin films of antimony as a function of thickness expressed in micrograms per cm² has been determined for blue, red, and white illumination. An abrupt change in optical properties of the antimony layer has been noted at the phase change. The transition occurs at a transmission of approximately 30 percent. Photosurfaces have been prepared by activating the deposited antimony films with cesium. The photoelectric yield for both regular and reverse illumination has been determined for different antimony thicknesses. For reverse illumination the peak response occurs between 5.5 and 6.0 micrograms per cm² of antimony. Expressing the photoresponse for reverse illumination of green light with decreasing transmission of the original antimony layer for blue, white, and red light shows peaks at 82, 88, and 92 percent, respectively. The transmission of the photosurface, which is substantially independent of processing or photoelectric efficiency of the surface, is given for blue, green, and white light.

IN the commercial practices of preparing semi-transparent photosurfaces, the thickness of the photosurface is determined and controlled by optical transmission measurements of the base metal,[1,2] e.g., antimony, silver, and bismuth-silver, for Cs–Sb, Ag–O–Cs, and Bi–Ag–O–Cs photosurfaces, respectively. The degree of oxidation in the two latter surfaces is quite often followed and controlled by observing the changes in optical transmission. An important factor, therefore, in the manufacture of uniform, maximum sensitivity photosurfaces is the correlation of the optical characteristics of the constituents with the photoelectric properties of this class of photosurfaces.

For a cathode in which the illumination is incident on the surface farthest from the collector (usually referred to as reverse illumination), thickness is critical because the combined paths of an absorbed quantum and the associated electron must traverse the entire thickness of the surface.

In this paper only the cesium-antimony surface will be considered.

EXPERIMENTAL

The tubes specially constructed for this investigation consisted of a mounted 25×75 mm microglass slide, a shielded antimony bead evaporator, and a source of cesium. Two longitudinal strips of platinum along the face of the slide were fired into the glass. Hanovium Liquid Platinum Bright No. 1[3] was used as the source of platinum. These strips served as electrical connections to the photosurface, and were contacted by inconel springs. All other metal parts and leads were made of nickel.

In the usual procedure a tube sealed onto the pumping position was baked at 375° to 400°C for 8 to 10 hours in a vacuum of the order of 10^{-6} mm Hg. The cesium source was then outgassed and antimony vaporized by resistive heating at temperatures below visible heat. Cesium was then released by eddy current heating a mixture of silicon and cesium dichromate. The surface was then sensitized at 100° to 180°C and the excess cesium pumped out of the tube.

A number of tubes were prepared with the antimony previously vaporized onto the slide at a vacuum pressure of 3×10^{-5} mm Hg in a demountable exhaust position. Antimony so deposited permitted greater evaporation to slide distances and, hence, a more gradual variation in antimony thickness. Tubes prepared with these pre-evaporated slides were baked at 270°C. Except for lower over-all sensitivity, these tubes gave the same results as those of the previous type.

Transmission was determined as the ratio of the light intensity transmitted through the film and glass support to the light transmitted, prior to the layer deposition, of the glass alone. The resultant values of transmission, therefore, do not take into account the reflectivity of the vaporized films. In the case of the completed photosurface, particularly at the higher thicknesses, this factor is of some significance.

Antimony transparency was measured in the mounted envelope for tubes of the first type, and for both the individual slide and the mounted tube for those of the second. Photosurface transmission was measured in the completed tube. A Photovolt 520M photometer and a tungsten lamp with and without glass filters were used for transmission measurements. The spectral response of the 2848°K tungsten lamp-filter combinations was determined by means of a G.E. Spectroradiometer, and the response of the 1P21 photomultiplier of the Photovolt was measured with a Beckman DU monochrometer and a calibrated tungsten light source.

Photoelectric yield was measured along the slides using a 2848°K tungsten lamp source with and without filters. The projected beam for reverse illumination on the slide was the width of the slide and about 3 mm high. Thickness of antimony layers was calculated from the amount of metal vaporized and the evaporation distances. The response of the lamp-filter combinations,

[1] A. Sommer, *Photoelectric Tubes* (Methuen and Company, Ld., London, 1951), second edition, Chap. III.
[2] V. K. Zworykin and E. G. Ramberg, *Photoelectricity and Its Application* (John Wiley and Sons, Inc., New York, 1949), Chap. V.
[3] Obtained from Hanovia Company, Newark, New Jersey.

indicated by the Corning designation, is given in Fig. 1 along with the sensitivity curve for the photomultiplier tube of the photometer. The response curves are brought to a common maximum. The true spectral curve for transmission is, therefore, the product of this last curve and that for a particular filter. The effect of the 1P21 response on the blue filter is negligible. However, in the case of the red filter, the peak of the effective curve is shifted toward the blue by about 20 mμ.

Points on the slide were characterized by the transmission of antimony; they were then correlated with the results of transmission of cesium-antimony after sensitization. The photoelectric yield for these points was then referred against the transmission of the photosurface. Therefore, though the method of determining the relation of antimony thickness (as expressed in $\mu g/cm^2$) to antimony transmission may be subject to considerable error, the interrelation of antimony transmission, photosurface transmission, and photoelectric yields was unaffected by any such errors.

RESULTS AND DISCUSSION

Beyond a certain thickness of antimony deposited at room temperature, it has been reported that the second modification of antimony "grows" with time from small circular areas until the entire surface is covered by the more opaque phase.[4,5] This apparent "growth" or crystallization occurs subsequent to the discontinuation of the antimony vaporization and lasts for several minutes.[6] In conjunction with this investigation, such behavior was observed on glass surfaces under certain conditions: when the vacuum was relatively poor (greater than 5×10^{-5} mm Hg), and the antimony was deposited onto an underbaked or insufficiently cleaned glass surface. The latter condition appeared to be the more critical. Under these circumstances the gradual phase transition was observed. The time necessary for the change depended upon the degree of contamination of the surface, increasing with the amount of surface impurity. The mottled circular pattern was never observed for a wedge of antimony vaporized on a clean surface. Instead, there was a clear line of demarcation dividing the two phases and tracing a path corresponding to a uniform thickness. This transition line was consistently straddled by values of optical transmission indicated at the break in Fig. 2. Although this sharp change was readily observable in the antimony layer, no such demarcation was visible after cesiation. Nor were there any marked corresponding changes in photoresponse at the phase boundary. Consequently, it is believed that the transition from the first modification of antimony (the metastable form) to the second form is very rapid on a clean glass surface at room temperature at the critical thickness.

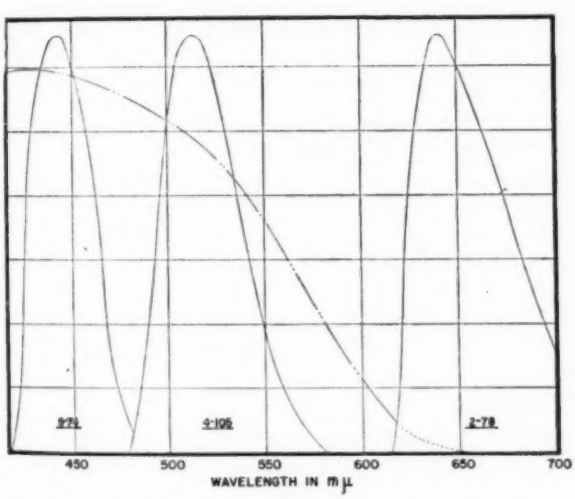


FIG. 1. Spectral responses of three narrow-band-pass Corning filters in combination with a 2848°K tungsten lamp, maximized at a common value. The broken line is the spectral sensitivity curve of the 1P21 photomultiplier of the detector.

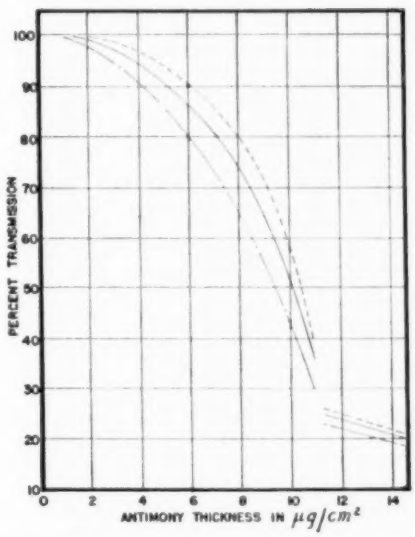


FIG. 2. Transmission of antimony in percent *vs* thickness expressed in $\mu g/cm^2$ for three sources illumination: ———— white light, – – – – blue (5–74), and — — — — red (2–78). The break in the curves at 11 $\mu g/cm^2$ is due to a phase change.

The variation of transmission with the color of the incident beam of light as indicated in Fig. 2 shows an increasing transparency of the antimony film from blue to red. For the second modification of antimony, the difference in transparency is considerably reduced. This is in close agreement with the observations of Soezima,[7] who found an abrupt change in optical characteristics corresponding to the change in phase. The variation of the visible light transmission with wavelength is so reduced with thicker layers that d'Or and Pirlot have suggested the use of antimony films as optically neutral filters at about 10 percent transmission.[8]

The relative transmission for the photosurface is

[4] H. Murmann, Z. Physik **54**, 755 (1929).
[5] R. Bowling Barnes and M. Czerny, Phys. Rev. **38**, 338 (1931).
[6] L. Harris, J. Appl. Phys. **17**, 757 (1946).
[7] Yosio Soezima, J. Sci. Research Inst. (Tokyo) **44**, 63 (1949).
[8] L. d'Or and G. Pirlot, Bull. soc. roy. sci. Liège **18**, 108 (1949).

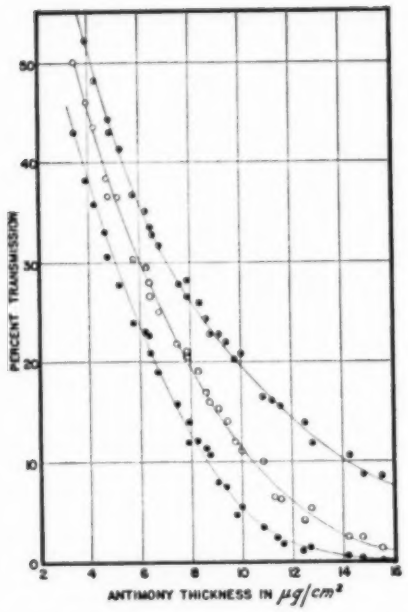


FIG. 3. Transmission in percent of Cs—Sb *vs* thickness of the original antimony layer expressed in $\mu g/cm^2$: ◑ = white light, ● = blue (5–74), and ○ = green (4–105).

given in Fig. 3 for blue, green, and white light. The overall relation is in close agreement with a similar plot of transmission of the cesium-antimony surface given by Dyatloviskaya.[9] Optical constants for this photosurface have been reported by Morgulis, *et al.*[10] Several of these values were later corrected by Dyatloviskaya.[9]

Figure 4 gives the photoelectric yield for reverse illumination of blue, green, and white light as a function of the original antimony thickness. Of particular interest are the relative slopes below and above the optimum

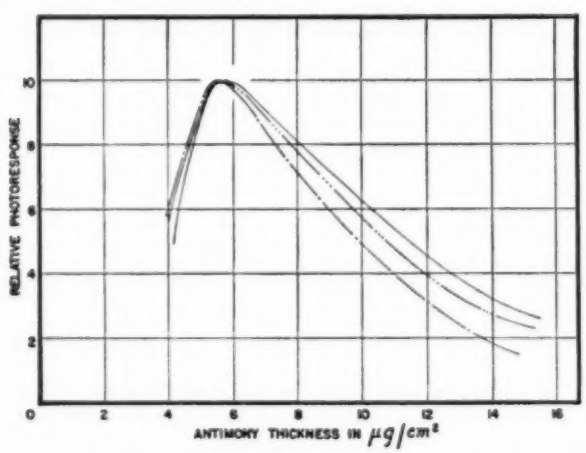


FIG. 4. Relative photoresponse of Cs—Sb under reverse illumination as a function of thickness of the original antimony layer expressed in $\mu g/cm^2$ for three sources of illumination. The curves are maximized to a common value: ———— white light, — — — — blue (5–74), and — — — — — green (4–105).

[9] B. I. Dyatloviskaya, J. Tech. Phys. (U.S.S.R.) **22**, 84 (1952).
[10] Morgulis, Borzyak, and Dyatloviskaya, Izvest. Akad. Nauk S.S.S.R. Ser. Fiz. **12**, 126 (1948).

thickness. For illumination of lower wavelength, the response drops more rapidly as the thickness of the surface is increased beyond the optimum thickness. However, decreasing the photosurface thickness in the region below the peak results in a more gradual drop in response for the blue as compared to higher wavelengths. This is in keeping with expectations based on the spectral absorption for cesium-antimony, which peaks in the blue and drops toward the red. For reverse illumination in which a released electron must travel $(d-x)$, where d is the thickness of the photosurface, and x is the depth at which a quantum is absorbed, increasing d results in a rapid diminution in response for $d \gg \bar{x}$; i.e., beyond the peak. For a given thickness of photosurface in this region (to the right of the maximum), the depth of photosurface which an electron must traverse, $(d-\bar{x})$, is less for longer wavelengths than for the blue because of the greater light absorption

TABLE I. Relative photoelectric yield of Cs—Sb for reverse illumination of three illuminants maximized at 10 and the corresponding transmissions of the antimony layer.

Percent transmission of antimony layer			Relative photoelectric yield		
(5–74)	White	(2–78)	(5–74)	(4–105)	White
91.4	95.0	97.1	6.10	5.75	4.50
89.0	93.2	96.0	7.62	7.40	6.82
86.5	91.3	94.6	9.14	9.10	8.81
81.8	88.0	92.1	9.96	10.0	10.0
80.4	86.7	91.0	9.70	9.86	9.91
77.1	84.0	88.8	9.12	9.35	9.48
73.3	81.1	86.3	8.42	8.81	9.00
69.4	77.7	83.5	7.70	8.26	8.55
65.2	73.9	80.3	7.05	7.82	8.08
60.4	69.4	76.3	6.45	7.18	7.60
55.0	64.3	71.5	5.87	6.65	7.12
49.0	57.8	65.6	5.32	6.13	6.77
42.5	51.5	58.1	4.81	5.65	6.20
38.7	47.0	52.7	4.54	5.35	5.94
30.5	36.6	39.5	4.00	4.80	5.40
23.0	24.7	25.8	3.67	4.56	5.18
20.7	22.3	23.0	2.37	3.22	3.83
18.5	19.9	20.6	1.50	2.40	2.85

in the latter. At thicknesses below the peak where $d < \bar{x}$, this higher absorption for the blue results in a consequent greater relative emission as compared to that for higher wavelength. An analysis of the optical dependence of photoemission of cesium-antimony in which both the incident quantum and the released photoelectrons are assumed to follow the logarithmic absorption law (single encounter) is given by Morgulis.[11] In this treatment the reflectance of the surface was not considered which may account for the anomalous results reported for regular illumination. In a later paper,[10] however, optical interference for regular illumination at higher wavelengths was considered.

In Table I are listed the percentage transmission for white, blue, and red light of the original antimony layer and the corresponding photoelectric yield of the re-

[11] N. D. Morgulis, Compt. rend. acad. sci. U.R.S.S. **52**, 675 (1946).

sultant photosurface with white, blue (5–74), and green (4–105) illumination. An examination of this table reveals the following information: the maximum photoelectric yield for both green and white reverse illumination corresponds to a thickness for which the transmission of blue (5–74), red (2–78), and white light of the original antimony layer is 81.8 percent, 92 percent, and 88 percent, respectively.

For irradiation with blue light the corresponding transmissions are slightly greater. Also, it may be seen that a variation in transmission with white light from 95 percent to 88 percent for the antimony layer results in a change in relative response from 6.10 to about 10 for the resultant photosurface under blue (5–74) light radiation, whereas under white light a considerably greater variation (4.50 to 10) is found. In addition, with the tabulated data, such questions as: "How great is the variation in response on a cathode

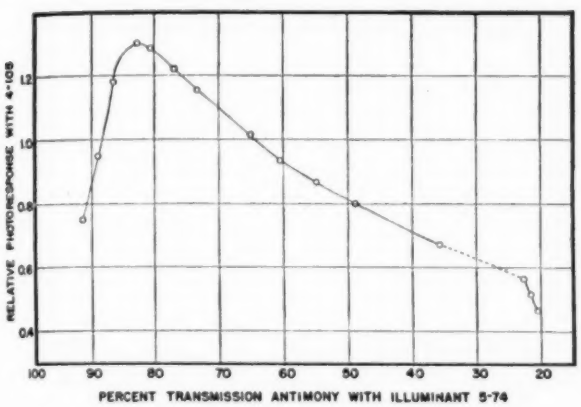


FIG. 5. Photoresponse in arbitrary units of Cs—Sb under reverse illumination *vs* transmission in percent of the original antimony layer. Photosurface illuminated with 4–105. Antimony transmission measured with illuminant 5–74.

which has a given fluctuation in antimony transmission?" can be quantitatively answered.

From Table I, then, can be evaluated the nine functions of photoelectric yield with transmission of the antimony base layer. One of these relations, photoresponse of 4–105 with antimony transparency of 5–74, is plotted in Fig. 5. (The ordinate is not maximized at 10 as in Table I.) The steeply rising portion of the curve at high transmission values indicates the poor uniformity of photoresponse with small variations of antimony transparencies in this region. Shifting the photosurface illumination toward the blue decreases the slope of this portion of the curve, whereas a shift toward the red increases the slope. The reverse effect is noted in the region of negative slope; shifting the irradiating wavelength to the blue causes a steeper slope, while a shift to the red results in a somewhat less severe drop. Changing the color of the light used in the transmission measurements of the antimony film from blue to red results in the following changes: (a) as noted above,

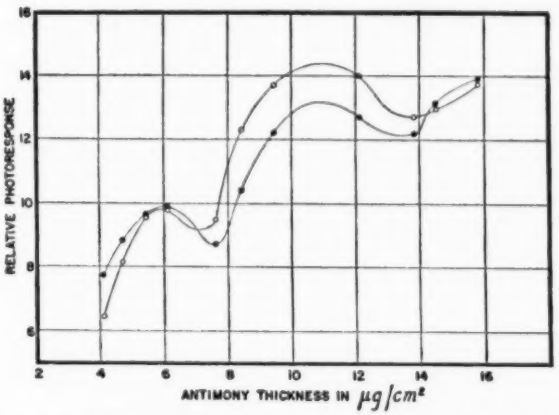


FIG. 6. Relative photoresponse of Cs—Sb with forward illumination *vs* thickness of the original antimony layer expressed in $\mu g/cm^2$. Illumination: ○=5–74 and ●=4–105.

the peak response is shifted from 82 percent to 92 percent transmission, (b) the slope of response with percentage transmission in the ascending region is considerably steeper, and (c) the slope of response with percentage transmission in the descending region is also somewhat steeper. It is evident, therefore, that a consideration of the color of light used in the transmission measurements of the antimony film, as well as in the illumination of the photosurface, is necessary in order to control photoresponse variations of a cathode.

Khorosh[12] has reported that the maximum sensitivity in $\mu a/lu$ for a Cs—Sb photocathode occurs at a thickness corresponding to a white light absorption of 50–60 percent in the antimony layer. Since these photosurfaces were reportedly prepared as individual cells of uniform antimony thickness from 10 to 90 percent light absorption, it is believed that the frequent large fluctuation of response from tube to tube may in part be responsible for this reported transmission value of antimony at the peak. In this investigation, the tube-to-tube fluctuation was obviated by employing a wedge. Under these conditions various thicknesses were treated under identical conditions.

In Fig. 6 is plotted the function of photoresponse for forward illumination (5–74 and 4–105) with antimony thickness expressed in $\mu g/cm^2$. The ordinate scale is in the same relative units as in Fig. 2, with the thickness of the peak reverse illumination response set at about 10. It can be seen that at greater thicknesses the curve is not a sensitive function of thickness and approaches the behavior of an opaque surface. The shape of this curve has been explained[9] in terms of reflectivity of the surface which shows a counter-phase relation with response for forward illumination. Therefore, a plot of forward illumination as a function of light *absorption* rather than transmission would be expected to yield a smoothly rising curve without changes in sign of the slope.

[12] D. M. Khorosh, J. Tech. Phys. (U.S.S.R.) **17**, 341 (1947).

Spin-Echo Memory Device*

SIDNEY FERNBACH† AND WARREN G. PROCTOR‡
Stanford University, Stanford, California
(Received March 6, 1954)

A proton-rich sample placed in a strong inhomogeneous magnetic field of mean strength H_0 was subjected to a pattern of relatively weak radio-frequency pulses at the Larmor frequency of the protons in the field H_0. The pattern was then recalled by applying a strong rf pulse at a later time as in the spin-echo technique. It is shown both mathematically and experimentally that such a series of pulses, varying in amplitude can be "memorized" by the spin system of protons for times as long as one second and then repeated, preserving both shape and relative amplitude.

I. INTRODUCTION

THE phenomenon of spin echoes, first observed by Hahn,[1] has been investigated as a possible basis for a memory device. It was decided to make the initial study as simple as necessary to demonstrate the feasibility of such a device. Therefore, the study was limited to the observation of the response of the memory system to a pattern of rectangular pulses. Furthermore, in order to minimize the experimental task, it was planned to study a sample of protons as a "memory unit," only for information which could be described by audio frequencies.

In this type of memory device, a material rich in protons—water, for example—is placed in a strong, inhomogeneous field of mean strength H_0. A transmitter is modulated to provide a pattern, or code, of a relatively weak magnetic field at the Larmor frequency of the protons in the field H_0. If the code is desired to be recalled at a time $t=2\tau$, then an intense pulse, called the reading pulse, is supplied at the time $t=\tau$. The recalled code is manifested as a component of nuclear rf magnetization in a direction perpendicular to H_0. A coil, located as in the method of nuclear induction,[2] has voltages induced in it due to the flux changes through it, and these voltages are amplified and detected. These detected voltages, corresponding to the envelope of the rf induced voltage, should correspond to the envelope of the transmitted code. The sequence of events is shown in Fig. 1, where the wave forms represent rf envelopes. It will be noted that the time sequence of the code has been reversed in the echo; it is also possible to obtain the proper sequence under slightly different conditions, described in Sec. II. Section III describes the apparatus used for these experiments, and Sec. IV presents the experimental results.

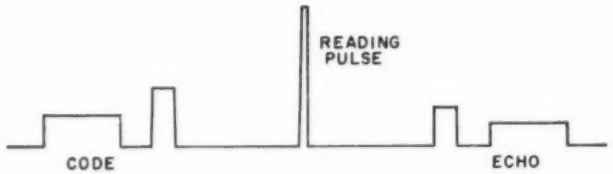


FIG. 1. Sequence of events in the spin-echo memory system.

II. MATHEMATICAL TREATMENT

The conditions under which a code can be reproduced at a later time are derived simply from the Bloch[3] differential equations for a nuclear induction system. In the notation of Bloch, with exceptions noted below, these equations are

$$\begin{aligned}(du)/(dt)+Sv&=-u/T_2\\(dv)/(dt)-Su+R_1w&=-v/T_2\\(dw)/(dt)-R_1v&=-(w-w_0)/T_1,\end{aligned}\tag{1}$$

where

$$R_1=\gamma H_1$$
$$S=\gamma\Delta H=\gamma H_0-\omega$$

and

$$w=M_z.$$

These equations are solved for the two intervals, (1) during a pulse, and (2) between pulses; then the solutions are matched at the boundaries of the intervals. The details of the calculations appear in the Appendix. It is shown there that if the transverse and longitudinal relaxation times T_2 and T_1 are equal ($=T$) and if diffusion effects are neglected, in the simple case of two pulses (represented by the subscripts 1 and 2 in the equations below) separated in time by τ, the following terms in v are responsible for the echo at time 2τ:

$$\frac{R_1\omega_0}{2}\left[(1-\cos\beta_2\phi)\frac{\sin\beta_1\theta}{\beta_1}+S^2\frac{\sin\beta_1\theta}{\beta_1}\frac{\cos\beta_2\phi-1}{2}\right]\times\cos S(t-2\tau-\phi+\theta)$$
$$-\frac{R_1S\omega_0}{2}\left[(1-\cos\beta_2\phi)\frac{\cos\beta_1\theta-1}{\beta_1^2}+S^2\frac{\cos\beta_1\theta-1}{\beta_1^2}\frac{\cos\beta_2\phi-1}{\beta_2^2}\right]\sin S(t-2\tau-\phi+\theta),\tag{2}$$

---

* Work done under contract with the U. S. Office of Naval Research 225(04).

† Now at University of California Radiation Laboratory, Livermore, California.

‡ Now at University of Washington, Seattle, Washington.

[1] E. Hahn, Phys. Rev. **80**, 580 (1950).

[2] Bloch, Hansen, and Packard, Phys. Rev. **70**, 455 (1946).

[3] F. Bloch, Phys. Rev. **70**, 460 (1946).

where

$$\beta_1^2=R_1^2+S^2, \quad \beta_2^2=R_2^2+S^2,$$

and the times are indicated in Fig. 2.

To get the total response near $t=2\tau$, these terms must be integrated over all S. It turns out that an exact reproduction of the input pulse is obtained when the condition $R_2 \gg S \gg R_1$ is imposed in the expression above. This then yields

$$-\frac{R_1\omega_0}{2}(1-\cos R_2\phi) \times\left[\frac{\sin S(t-2\tau+\phi)}{S}-\frac{\sin S(t-2\tau-\phi+\theta)}{S}\right], \quad (3)$$

which when integrated over all S results in a square pulse with leading edge at $t=2\tau+\phi-\theta$ and trailing edge at $t=2\tau+\phi$. The amplitude is a maximum when $R_2\phi=\gamma H_2\phi=\pi$.

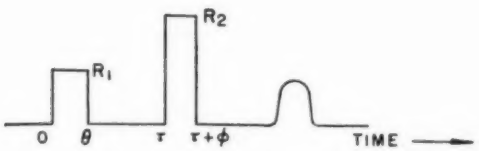


FIG. 2. Nomenclature in the case of a single code pulse R_1, and reading pulse R_2.

Thus the conditions for good reproduction may be summarized briefly as

$$\begin{aligned} &(a)\ S_{\max}\theta \gg 1 \\ &(b)\ R_2\phi=\gamma H_2\phi \cong \pi \\ &(c)\ R_1 \ll S_{\max}. \end{aligned} \quad (4)$$

More specifically:

(a) The inhomogeneity of the magnetic field over the sample should be large enough to provide a band width or range of nuclear Larmor frequencies sufficient to accommodate all the frequencies contained in the modulated code pulses.

(b) The reading pulse should be sufficiently intense to cause all nuclei in the sample to rotate about the direction of its field by π radians.

(c) The deviation of any magnetization from the direction of the dc field caused by the code pulses should be small, in order that the nuclear response may be linear in amplitude or duration. For example, these conditions, applied to our initial program, made the following requirements. Since audio information up to 20 kc was to be stored, the variation of dc field over the sample, measured from the mean field strength, had to be ± 20 kc/$\gamma=\pm 5$ gauss (the gyromagnetic ratio γ for protons being 4.25 kc/gauss). Code pulses were not to be shorter than 1/20 kc$=50$ μsec. The reading pulse intensity was to be somewhat greater than 5 gauss, and the intensity of the code pulses considerably smaller, particularly if the code were to be of some duration. In Fig. 3 is shown the echo response calculated for various choices of $S_{\max}$, assuming the conditions above and $\theta=200$ μsec.

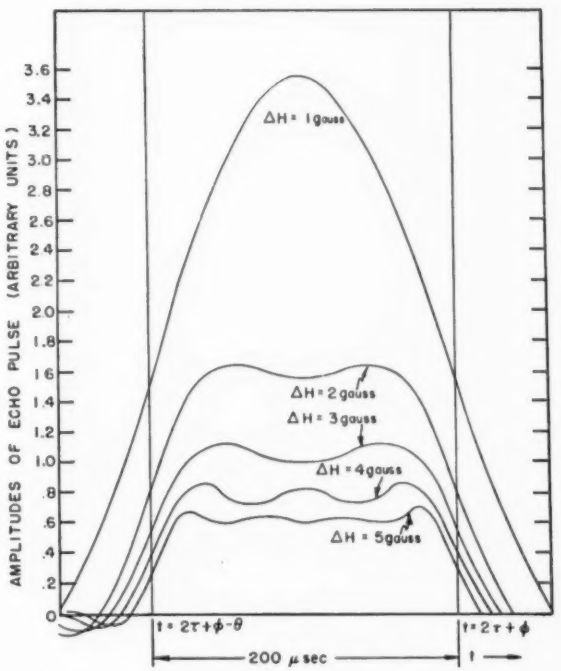


FIG. 3. Calculated response from a square pulse input for several values of the field gradient. Ordinate in arbitrary units.

The calculations representing a more general case show that under the conditions above, spacing between and during pulses is preserved and that the amplitudes of the echoes are in proportion to the amplitudes of the input pulses. The echoes, however, appear in reverse order, each being as far from the trailing edge of the reading pulse as its original is from the leading edge of the reading pulse. Also it is important to note that no echoes other than the desired reproductions appear if the above conditions are met. It is possible to invert the order by applying another reading pulse at a later time, this new pulse reversing all before it.

Although T_1 was assumed equal to T_2, it is clear that the memory time is limited by T_2 (or possibly diffusion effects if these are great). That this is so, can be seen from the fact that the echo decays like $\exp(-t/T_2-k\tau^3/3)$, where k depends upon the field gradient and the self-diffusion coefficient of the spin-containing molecule and $t \cong 2\tau$.

There is another procedure, depending upon the phenomenon Hahn[1] called the stimulated echo, for obtaining echoes for times as great as T_1, rather than T_2. This method has the advantage, in addition to the fact that $T_1 \geqq T_2$, that the echoes appear in proper sequence immediately after the reading pulse. In this case a strong pulse (the starting pulse), is applied to the spin system before the code, and the rest of the pulses follow the same sequence as before. Each of the strong pulses now flips the spin systems by $\pi/2$ instead of π. This

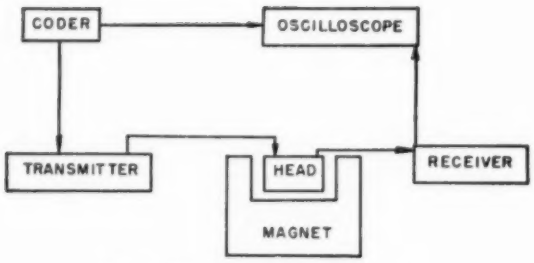


FIG. 4. Block diagram of apparatus.

results in the stimulated echoes of the pulses, spaced so that each appears as far behind the reading pulse as its original appearance behind the starting pulse. The mathematical treatment closely parallels that for the previous case; it is significant, however, that the echoes decay like $\exp(-t/T_1)$, and hence memory time depends on T_1 rather than T_2. Of course, the information code has to be inserted in a time small compared to T_2, hence the capacity depends on T_2.

III. APPARATUS

The apparatus required to observe spin echoes is indicated by the block diagram of Fig. 4. The coder is an assembly of interconnected multivibrators which provide pulses used to gate the transmitter. The coder also provides a synchronizing pulse to the oscilloscope, to start its sweep. The head contains the sample and rf coils. One of these, the transmitter coil, is excited by the transmitter and provides flux for the sample region. The other, the receiver coil, is approximately at right angles to the transmitter coil and therefore not coupled strongly to it; the receiver coil surrounds the sample and has induced in it rf voltages from the precessing nuclei. The voltages are amplified and detected by the receiver and appear on the oscilloscope. The magnet, of course, provides a strong constant field over the sample region, a field which, as has been pointed out, must possess a desired degree of inhomogeneity. These various components will not be described in detail, since their design and construction is straightforward or, as in the case of the permanent magnet which was used in these experiments, a detailed description would be excessively lengthy. However, remarks concerning details of interest about these components are gathered in the pertinent paragraphs below.

(a) The Transmitter

The transmitter used was powered by a 4E27 final stage; a 6SN7 7.5-Mc Hartley oscillator excited a 6AG7 doubler, which in turn excited an 807 doubler driver. The tank coil of the oscillator was damped by the cathode impedance of the second section of the 6SN7, which damping was removed by negative gates applied to its grid. This provided 7.5-Mc oscillations of equal amplitudes for both the code and reading pulses. The amplitude of the transmitted pulses was controlled by modulating the screen grid of the 807 by positive gates, which raised the excitation to the final stage. All stages were class C, using power only when excitation was present.

It was not necessary to achieve fast rise and fall times in the transmitter, hence resistive loading of tuned circuits was not necessary for the purposes of these experiments. The resonant tank circuit was coupled to the tuned transmitter coil in the head by capacitively dividing each to a point of low impedance and connecting these points by a cable.

(b) The Head

The head consisted, first, of a 5-turn coil, wrapped about a form which would accommodate the $\frac{1}{2}$-in. o.d. test tubes used as sample holders. This coil, the receiver coil, was resonated to 30 Mc by the capacitance of the short cable connecting it to the receiver, plus a small variable capacitor located in the receiver. The receiver coil was located between and perpendicular to two single loops 2.9 cm in diameter, spaced 2 cm apart, and connected in series, which together formed the transmitter coil. It will be recognized that the spacing is more than that required by Helmholtz dimensions, but calculations showed that the flux intensity could be considerably increased while impairing the homogeneity of the rf flux very little. As described by Weaver[4] the position of one of the loops was made adjustable by means of a screw which moved the Lucite block upon which it was wrapped. This allowed the coupling from the transmitter to the receiver to be minimized. The coupling was further reduced by using a u-mode paddle, also described by Weaver. Here, in contrast to the ordinary method of nuclear induction, careful isolation of the receiver from the transmitter is not necessary because the nuclear response only occurs after the transmitter is quiet. The purpose of the balancing, as it is called, is only to protect the receiver and prevent excessive blocking. As stated above, the transmitter coil was resonated by a capacitor located in the head and driven at a point of low impedance. 160 watts of rf power was necessary to provide about 6 gauss H_1 over the sample volume, which was about 3 cc.

(c) The Receiver

The receiver used was a war surplus APS-15 IF strip, with only minor modifications. One of these, as mentioned, was the input circuit, which was arranged so that the grid of the first stage amplified what voltage was present across the resonating capacitor of the receiver coil. The video stages were dc connected, because of the long time intervals sometimes used between pulses.

Contrary to the practice in observing conventional nuclear induction signals, there is no homodyning voltage of large amplitude present at the detector.

[4] H. E. Weaver, Jr., Phys. Rev. 89, 923 (1953).

Consequently, the detection was not linear for very small signals, but the situation was alleviated by the high gain of the rf amplification.

(d) The Magnet

A permanent magnet, using 1040 lbs of Alnico V and $1\frac{1}{2}$ tons of yoke steel, was built to provide a well stabilized field of 7165 gauss between pole faces $7\frac{3}{4}$ in. in diameter and $1\frac{3}{4}$ in. spacing. The large gap volume and the high flux intensity were responsible for the amount of Alnico V needed. The magnet was designed with generous dimensions to allow for experiments of other kinds, in which homogeneous fields might sometimes be necessary or large sample volumes used. Uniform and controlled inhomogeneities were generated by tipping one or both poles, using brass jacks placed between the poles and the yoke, so that the pole faces were no longer parallel.

(e) The Coder

The coder was made up of 15 6SN7 double triodes, used as delay or gate multivibrators, cathode followers, etc. This instrument provided the positive and negative square gates for modulating the transmitter in the condition described. Normally, up to three separate code pulses and one reading pulse were available.

It might be appropriate, in conclusion, to make some remarks about the relationship between the rf power requirements and the signal intensity and frequency response. Let us assume that a satisfactory memory device of band width f_0 has been made which has a transmitter delivering power P_0 during the reading pulse of flux intensity H_{10} to a sample of volume V_0; we further assume that f_0 is so large that resonant coils must be damped. The question as to how much more power is necessary in order to store information with band width Nf_0 with the same signal-to-noise ratio available from the receiver then can be determined as follows: Since the desired band width is equal to Nf_0, the field inhomogeneity must be increased to a new value $N\Delta H_0$. This reduces the number of nuclei per unit frequency interval, or field interval, to a fraction $1/N$ of its previous number, so that it becomes necessary to make $V=NV_0$. The transmitter must now supply a flux intensity NH_{10} over a volume NV_0 using a coil with $Q=Q_0/N$. Since the radio-frequency power delivered to a coil proportional to (flux intensity)^2 (volume) Q^{-1}, we now have that $P=N^4P_0$. However, the increased band width now admits N times more noise power. Hence, conserving signal-to-noise ratio, these crude considerations show that the radio-frequency power goes up as the 4.5 power of the band width, at high band widths.

IV. EXPERIMENTAL RESULTS

Figure 5 is a photograph of the oscilloscope trace showing a typical response of the system. The complete time sequence of the events, time progressing from right to left, is shown in the upper trace. This trace should be compared to Fig. 1. (The code and reading pulses both saturate the receiver, hence they appear as breaks in base line.) The center trace shows the envelope of the input code rf pulses, with the receiver gain lowered; the lower trace shows the nuclear response. The upper trace has a sweep speed of 300 μsec per centimeter, and the lower two a speed of 60 μsec/cm. The reticule is marked off in centimeters and appears reduced in the photograph. Here the delay, or memory time, is only 2.0 milliseconds and was kept purposely short in order to make the upper trace show some of the features of the code and its echo as well as the sequence as a whole. Since the sample material was glycerine, for which $T_2\cong 20$ milliseconds, the delay could have been made $\sim$10 msec without serious attenuation of the echo. The repetition rate used was 3.0 cps, or a repetition period of 330 msec. The repetition period must always be long compared to T_1, else unwanted small stimulated echoes appear. T_1 for glycerine is $\sim$37 msec; a repetition rate of 6 cps could have been used quite as satisfactorily.

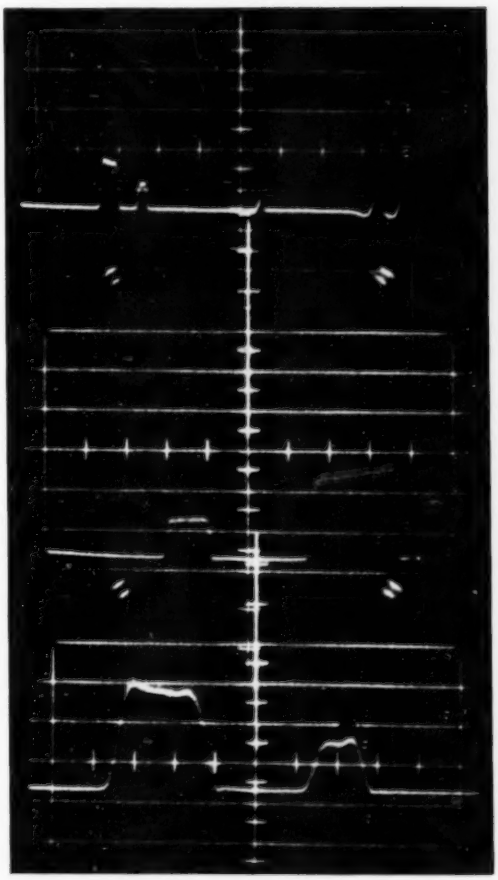


FIG. 5. Oscilloscope trace showing response to a complex code: (upper) the complete sequence, (center) the code, (lower) the echo. (Time reads from right to left.)

It will be noted that the rise and fall times of the

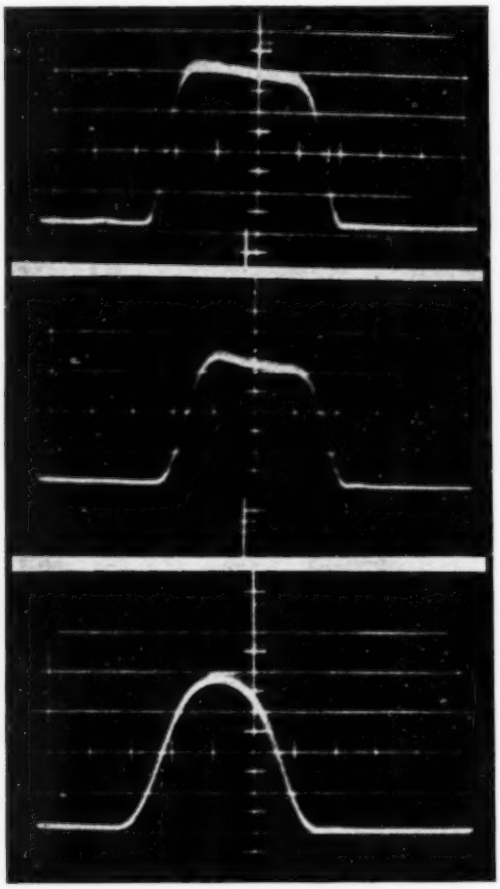


FIG. 6. Effect on echo of decreasing the field inhomogeneity: (upper) $\Delta H_{max}=4.9$ gauss, (center) $\Delta H_{max}=3.1$ gauss, (lower) $\Delta H_{max}=1.7$ gauss.

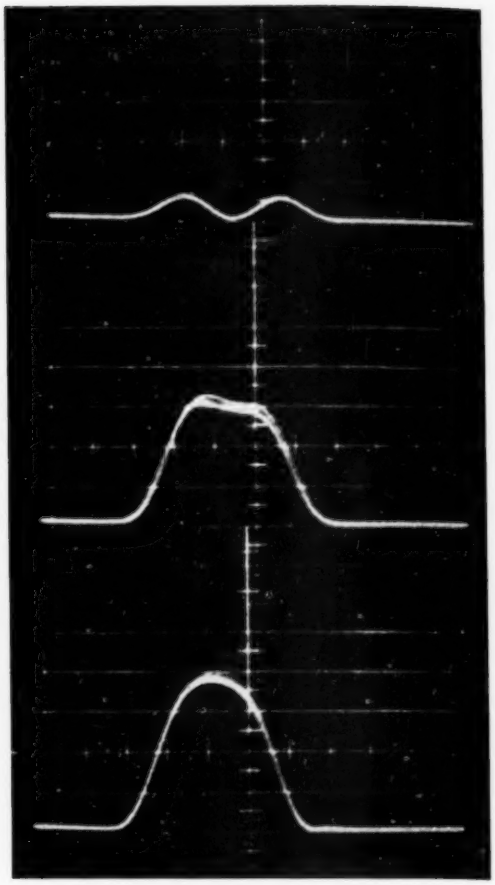


FIG. 7. Effect on echo of centering transmitter frequency with respect to the sample Larmor frequencies: (upper) poor centering, (center) transmitter frequency slightly off, (lower) centered.

echoes are about 30 μsec. This is about what one would expect from a value $H_2 \cong 5$ gauss and $\Delta H \cong 5$ gauss, as seen from Fig. 3. The measured gradient for this example was 13.1 gauss/cm. Since the sample was cylindrical, of 10.5-mm diameter, the field gradient being imposed across the cylinder, rather than along it, the total variation of field across the sample was 13.8 gauss. The number of nuclei at each field value, when plotted, would be proportional to the amplitude of a semicircle, extending from $(H_0-6.9)$ gauss to $(H_0+6.9)$ gauss. Replacing the circular cross section by that of the inscribed square, we may then say that ΔH_{max} is approximately $6.9/\sqrt{2}=4.9$ gauss. Although we did not experimentally provide samples with uniform proton densities per unit field, the experimental results compare well with the calculated performances shown in Fig. 3.

For a given H_2, it turns out experimentally that the frequency response improves with increasing ΔH_{max}, until $\Delta H_{max} \cong H_2$. For larger ΔH_{max}, parts of the sample are not being utilized. The theoretical analysis has shown that accurate reproduction of the code pulses may be expected when $H_2 \gg \Delta H_{max}$. Consequently, using a field inhomogeneity which causes $\Delta H_{max} > H_2$, response comes principally from those parts of the sample for which $H_2 \gg \Delta H$. In those parts of the sample where $H_2 \cong \Delta H$, some small high-frequency contribution may be expected; this case was not calculated because of its complexity. No contribution is expected from those parts of the sample for which $\Delta H \gg H_2$.

These considerations are illustrated in the traces of Fig. 6. For each of these, $H_2=3.5g$; ΔH_{max}, computed as above, is 4.9, 3.1, and 1.7 gauss, respectively, reading from top to bottom. These responses should be compared to curves which may be interpolated between those given in Fig. 3.

Figure 7 illustrates as well the effect of the location of the transmitter frequency with respect to the distribution of Larmor frequencies in the sample. In the lower trace, the transmitter frequency is centered with respect to the sample Larmor frequencies. In the center trace, the frequency has been moved partially to one side of the Larmor frequency distribution. This has the effect of increasing ΔH_{max}, at least in one "side band." When carried too far the improvement in signal shape, noted in the center trace, is lost, and only high-

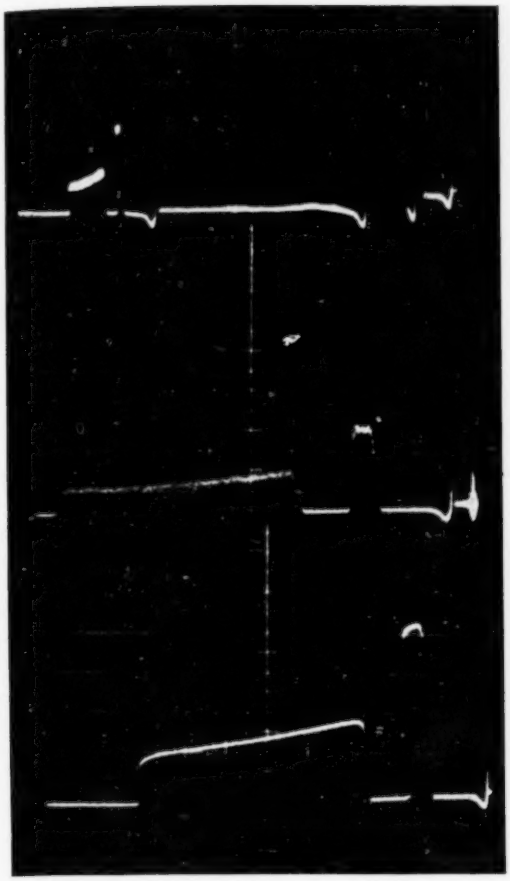


FIG. 8. Complex spin echo pattern resulting from use of starting and reading pulses: (upper) complete sequence, (center) the echo, (lower) the code. (Time reads from right to left.)

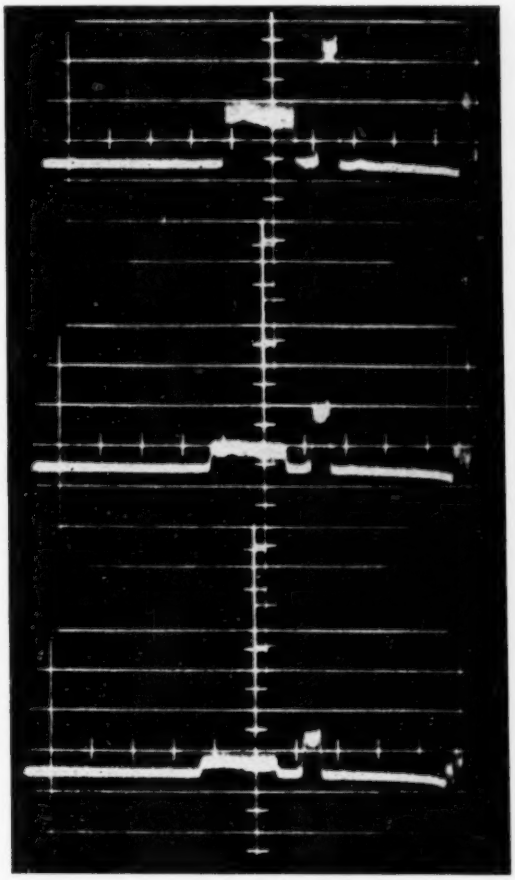


FIG. 9. Effect of relaxation time upon echo: (upper) immediate echo, (center) 0.3-sec delay, (lower) 0.6-sec delay.

frequency components of the signal are available, as seen in the upper trace.

Figure 8 shows operation when using a starting pulse as well as a reading pulse; here the echo, as observed, should show the same time sequence of pulses as in the code. The upper trace (500 μsec/cm) shows the complete cycle of events, time reading from right to left, while the center trace (100 μsec/cm) is an expanded sweep trace of the echo, to be compared with the input code in the lower trace (100 μsec/cm). Here $H_2 \cong \Delta H_{\max} \cong 5$ g.

As pointed out in Sec. II, the attenuation in this case is given by T_1, rather than T_2, provided that the code follows the starting pulse in a time short compared to T_2, and that diffusion effects are negligible. The upper trace of Fig. 9 shows an immediate echo; the center trace shows the response after a delay of 0.3 second, while the lower trace shows the response after a memory time of 0.6 second. Here all sweeps speeds are 50 μsec/cm, and $H_2 \cong \Delta H_{\max} \cong 6$ gauss. The sample used was Wildroot Cream-Oil, for which $T_2 \cong 0.3$ second and $T_1 \cong 1.0$ second at this value of field inhomogeneity. Note that the response is not distorted, but simply attenuated as the delay is increased. This is to be expected since the number of nuclei at each field value is attenuated in proportion to that number.

Figure 10 shows the response of the apparatus to a series of about 35 pulses each 20 μsec wide and 30 μsec apart. The upper trace (500 μsec/cm) shows the complete cycle, while the lower traces are expanded sweeps (100 μsec/cm) of the input pulses and the echo. Note that the ninth pulse is somewhat larger than the others in the input, and that this feature is reproduced in the echo.

Some investigations have been made seeking a suitable sample substance. Samples may be either gaseous, liquid, or solid. The gaseous samples were not examined because of their low densities. Solid samples may exhibit long thermal relaxation times, several minutes long,[5] but the transverse relaxation times are generally of the order of 10 μsec[1] for protons. This order of magnitude for T_2 can be easily estimated by considering how strong a field one nucleus produces at the position of another when separated by atomic distances. Because of this small T_2, solids were not investigated.

In liquids, because molecules rotate and change

[5] See, for example, R. V. Pound, Phys. Rev. 81, 156 (1951).

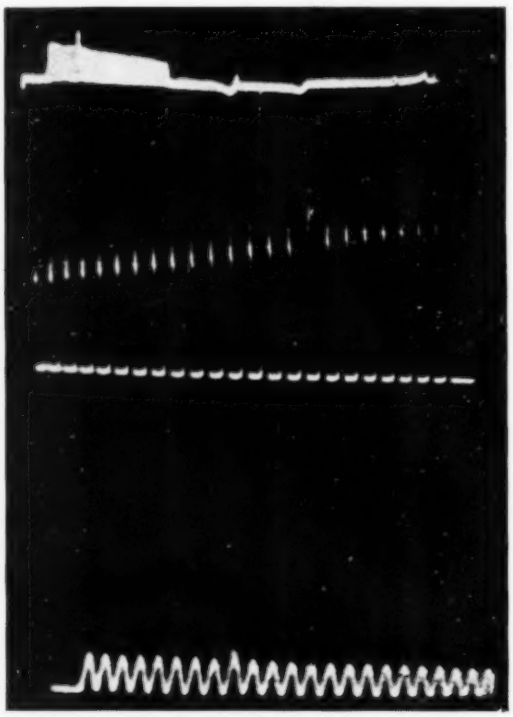


FIG. 10. Response to a code consisting of many pulses: (upper) complete sequence, (center) the code, (lower) the echo. (Time reads from right to left.)

positions relative to each other by significant amounts in times short compared to the Larmor frequency, such perturbing fields are largely cancelled out, so that both T_1 and T_2 may be long.[6] The greater the macroscopic viscosity of liquid, the more like a solid it appears, and the shorter T_2 becomes. In general, both T_1 and T_2 increase with the fluidity of the liquid. Hence, at room temperatures, water is probably the most suitable substance to use for long memory times, except for its rapid self diffusion. Diffusion of the molecules from one part of the sample to another during the memory period destroys the echo since the protons in that molecule have a different Larmor period than that which they had at the beginning of the period. When higher frequency response is desired, larger field gradients must be used, and the diffusion restriction is much more stringent. Diffusion enters the equations in a manner similar to T_2, as shown by Hahn.

These considerations lead toward seeking a sample substance which is highly fluid, but in which the molecules are constrained within tiny cells. Several mechanical methods for achieving this were tried without success. The use of emulsions was suggested as a chemical means of performing the same end and proved moderately successful. In an emulsion, the molecules within a globule, as well as those captured between globules, are confined by the globule walls and act as if in a pure liquid, but the globule itself is too large to migrate. Water-oil emulsions, made in this laboratory, proved to be highly fluid, with varying globule sizes, and were not satisfactory. An oil-lanolin emulsion, as is used commercially in cosmetics, has particle sizes, as examined under a microscope, uniformly near 1 μ in diameter. In fields only slightly inhomogeneous, relaxation times measure up to 1.2 seconds, which is not far short of the value 2.3 seconds for pure water.

Lanolin, one of the components of the above emulsions will take up water to the extent of 200 percent of its own weight in the anhydrous form. The resulting substance, chemically known as a mixture, was investigated because of the promising results shown by the emulsions. Lanolin itself has short relaxation times. After it takes up water however, two relaxation times become evident, a short one due to the lanolin and a long one, presumably due to the water. The more water held in the mixture the greater the importance of this longer relaxation time, with the result that the over-all relaxation time of the mixture is lengthened. The lanolin in the emulsions also contains water, so that the good results obtained with such samples may in part be due to the hydrous formation.

A few of the materials investigated are listed in Table I; some unusual materials are included since they might prove useful in specific applications or because they are instructive examples. No attempt has been made to measure the actual relaxation times in liquids; the values listed are what we call apparent relaxation times, or perhaps, useful memory times, and are simply taken from the average slope of the major portion of the decay curve when plotted on semilog paper. The decay curves for many liquids are not linear, but illustrate the strong influence of diffusion. T_1 (apparent) was measured by the stimulated echo method and will not in general be completely free from influence of T_2 and diffusion. The field gradient appears

TABLE I. Apparent relaxation times and relative signal amplitudes of various substances.[a]

Substance	Field gradient in gauss/cm	T_2 apparent	T_1 apparent	Rel. sig. amp
Tap water	4.2	20 msec	···	1.00
Tap water	1.2	57 msec	···	···
Lanolin, anhydrous	2.0	8 msec	48 msec	···
Lanolin, sat. with H_2O	2.1	0.44 sec	1.4 sec	···
	10.2	0.25 sec	1.1 sec	1.2
	20.0	0.25 sec	0.8 sec	···
Wildroot Cream Oil	2.1	0.46 sec	1.2 sec	···
Wildroot Cream Oil	10.2	0.29 sec	0.9 sec	1.6
Wildroot Cream Oil	20.0	0.21 sec	0.7 sec	···
Burroughs-Wellcome Toilet Lanolin	2.0	0.4 sec	1.0 sec	1.7
Solid petrolatum	2.0	30 msec	80 msec	···
Liquid petrolatum	2.0	43 msec	58 msec	2.4
Ethylene glycol	2.0	60 msec	0.3 sec	···
glycerine	2.0	22 msec	37 msec	···
various rubbers	2.0	1–2 msec	30–60 msec	···
butter	2.0	68 msec	0.26 sec	···

[a] In view of the information presented in Table I, it is believed that oil-lanolin emulsions provide the best sample materials for long memory times.

[6] Bloembergen, Purcell, and Pound, Phys. Rev. **73**, 679 (1948).

in column two. The last column used in each measurement shows the relative echo amplitudes at a common, moderate delay.

V. ACKNOWLEDGMENT

We wish to express our gratitude to Professor F. E. Terman of the Electronics Research Laboratory, and Professor F. Bloch of the Physics Department at Stanford University, for the privilege of working on this project. We are indebted to them and Dr. E. Hahn for many helpful discussions. We also wish to thank Dr. R. E. Norberg for pointing out the possibility of using the stimulated echo technique to increase memory times.

APPENDIX

We start with Bloch's equations.[3] (We shall not take diffusion into account, with the hope that a sample material may be found in which diffusion effects are small.) We thus have

$$\begin{aligned}(du)/(dt)+Sv&=-u/T_2\\(dv)/(dt)-Su+R_1w&=-v/T_2\\(dw)/(dt)-R_1v&=-(w-w_0)/T_1,\end{aligned}\tag{I-1}$$

where $R_1=\gamma H_1$, $S=\gamma\Delta H$, $w=N_z$, and the other symbols are as in Bloch's paper.

We proceed as Hahn[1] has done in his calculations. We find the solution to Eq. (I-1) during a pulse, assuming R_1 is constant and pulse width negligibly small compared to relaxation times; and the solutions between pulses when $R_1=0$. During the pulse the solution is

$$\begin{aligned}u(t)&=\frac{S}{\beta}A[\cos(\beta t+\xi)-\cos(\beta t_i+\xi)]+u(t_i)\\v(t)&=A\sin(\beta t+\xi)\\w(t)&=-\frac{RA}{\beta}[\cos(\beta t+\xi)-\cos(\beta t_i+\xi)]+w(t_i),\end{aligned}\tag{I-2}$$

where $\beta^2=R^2+S^2$, and A, ξ, $u(t_i)$, and $w(t_i)$ are determined by initial conditions and by assuming $u(t_i)^2+v(t_i)^2+w(t_i)^2=u(t)^2+v(t)^2+w(t)^2$.

The solution in absence of pulses is

$$\begin{aligned}u(t)&=[u(t_i)\cos S(t-t_i)-v(t_i)\sin S(t-t_i)]\times\exp\left(-\frac{t-t_i}{T}\right)\\v(t)&=[v(t_i)\cos S(t-t_i)+u(t_i)\sin S(t-t_i)]\times\exp\left(-\frac{t-t_i}{T}\right)\\w(t)&=w_0+[w(t_i)-w_0]\exp\left(-\frac{t-t_i}{T}\right),\end{aligned}\tag{I-3}$$

where we have assumed $T_1=T_2=T$. Let us now consider the case represented in Fig. 2. We wish to find the v mode response for $t\geqq\tau'$. From (I-3),

$$v(t)=[v(\tau')\cos S(t-\tau')+u(\tau')\sin S(t-\tau')]\times\exp\left(-\frac{t-\tau'}{T}\right).$$

We can find $u,v,w(\tau')$ in terms of $u,v,w(\tau)$ from (I-2); then u, v, $w(\tau)$, in terms of θ from (I-3) again; thus $v(t)$ in terms of θ, τ, R_1, and R_2. Using (I-2) for the first pulse where $t_i=0$, we obtain

$$\begin{aligned}u(\theta)&=-\frac{R_1Sw_0}{\beta_1^2}(\cos\beta_1\theta-1)\\v(\theta)&=-\frac{R_1}{\beta_1}w_0\sin\beta_1\theta\\w(\theta)&=w_0\frac{R_1^2}{\beta_1^2}(\cos\beta_1\theta-1)+w_0.\end{aligned}\tag{I-4}$$

In the second pulse

$$\begin{aligned}u(\tau')&=\frac{S}{\beta_2}A[\cos(\beta_2\tau+\beta_2\phi+\xi)-\cos(\beta_2\tau+\xi)]+u(\tau)\\v(\tau')&=A\sin(\beta_2\tau+\beta_2\phi+\xi)\\w(\tau')&=-\frac{R_2A}{\beta_2}[\cos(\beta_2\tau+\beta_2\phi+\xi)-\cos(\beta_2\tau+\xi)]+w(\tau),\end{aligned}\tag{I-5}$$

where A and ξ can be determined from

$$A\sin(\beta_2\tau+\xi)=v(\tau)\tag{I-6}$$

and

$$u^2(\tau+\phi)+v^2(\tau+\phi)+w^2(\tau+\phi)=u^2(\tau)+v^2(\tau)+w^2(\tau).\tag{I-7}$$

We obtain from Eq. (I-7),

$$A\cos(\beta_2\tau+\xi)=\frac{S}{\beta_2}u(\tau)-\frac{R_2}{\beta_2}w(\tau).\tag{I-8}$$

Using (I-6) and (I-8) in (I-5), we obtain

$$\begin{aligned}u(\tau')&=\frac{S}{R_2w_0}[v(\phi)v(\tau)-u(\phi)u(\tau)]+\frac{u(\phi)w(\tau)}{w_0}+u(\tau)\\v(\tau')&=\frac{v(\phi)}{w_0}\left[w(\tau)-\frac{S}{R_2}u(\tau)\right]+v(\tau)\cos\beta_2\phi\\w(\tau')&=\frac{1}{w_0}[w(\phi)w(\tau)+u(\tau)u(\phi)-v(\tau)v(\phi)].\end{aligned}\tag{I-9}$$

Directly from (I-5),

$$u(\tau)=[u(\theta)\cos S(\tau-\theta)-v(\theta)\sin S(\tau-\theta)]\times\exp\left(-\frac{\tau-\theta}{T}\right)$$

$$v(\tau)=[v(\theta)\cos S(\tau-\theta)+u(\theta)\sin S(\tau-\theta)]\times\exp\left(-\frac{\tau-\theta}{T}\right) \quad \text{(I-10)}$$

$$w(\tau)=w_0+[w(\theta)-w_0]\exp\left(-\frac{\tau-\theta}{T}\right).$$

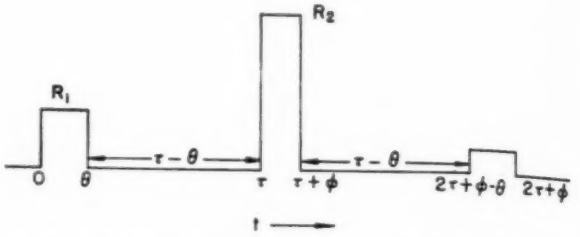


FIG. 11. Definition of quantities used in calculation for case of single code pulse and reading pulse.

From (I-4), and a similar equation for ϕ, (I-9), and (I-10) then

$$\begin{aligned}
v(t)=\exp\left(-\frac{t-\phi-\theta}{T}\right)\Bigg\{&-\frac{R_1w_0}{2}\Bigg[2S^2\frac{\cos\beta_1\theta-1}{\beta_1^2}\frac{\sin\beta_2\phi}{\beta_2}\\
&+(1+\cos\beta_2\phi)\frac{\sin\beta_1\theta}{\beta_1}+S^2\frac{\sin\beta_1\theta}{\beta_1}\frac{\cos2\phi-1}{\beta_2^2}\Bigg]\cos S(t-\phi-\theta)\\
&+\frac{R_1w_0}{2}\Bigg[(1-\cos\beta_2\phi)\frac{\sin\beta_1\theta}{\beta_1}+S^2\frac{\sin\beta_1\theta}{\beta_1}\frac{\cos\beta_2\phi-1}{\beta_2^2}\Bigg]\cos S(t-2\tau-\phi+\theta)\\
&-\frac{R_1Sw_0}{2}\Bigg[(1-\cos\beta_2\phi)\frac{\cos\beta_1\theta-1}{\beta_1^2}+S^2\frac{\cos\beta_1\theta-1}{\beta_1^2}\frac{\cos\beta_2\phi-1}{\beta_2^2}\Bigg]\sin S(t-2\tau-\phi+\theta)\\
&-\frac{R_1SW_0}{2}\Bigg[-2\frac{\sin\beta_1\theta}{\beta_1}\frac{\sin\beta_2\phi}{\beta_2}+(1+\cos\beta_2\phi)\frac{\cos\beta_1\theta-1}{\beta_1}+S^2\frac{\cos\beta_1\theta-1}{\beta_1^2}\frac{\cos\beta_2\phi-1}{\beta_2^2}\Bigg]\sin S(t-\phi-\theta)\\
&-R_1^2R_2w_0\frac{\cos\beta_1\theta-1}{\beta_1^2}\Bigg[\frac{\sin\beta_2\phi}{\beta_2}\cos S(t-\tau-\phi)+S\frac{\cos\beta_2\phi-1}{\beta_2^2}\sin S(t-\tau-\phi)\\
&-R_2\exp\left(\frac{\tau-\theta}{T}\right)\frac{\sin\beta_2\phi}{\beta_2}\cos S(t-\tau-\phi)+S\frac{\cos\beta_2\phi-1}{\beta_2^2}\sin S(t-\tau-\phi)\Bigg]\Bigg\}. \quad \text{(I-11)}
\end{aligned}$$

We are interested in the terms containing $\cos\sin S(t-2\tau-\phi+\theta)$, since these are the ones responsible for the echoes near $t=2\tau$. These are

$$\begin{aligned}
&\frac{R_1w_0}{2}\Bigg[(1-\cos\beta_2\phi)\frac{\sin\beta_1\theta}{\beta_1}+S^2\frac{\sin\beta_1\theta}{\beta_1}\frac{\cos\beta_2\phi-1}{\beta_2}\Bigg]\cos S(t-2\tau-\phi+\theta)\\
&-\frac{R_1Sw_0}{2}\Bigg[(1-\cos\beta_2\phi)\frac{\cos\beta_1\theta-1}{\beta_1^2}+S^2\frac{\cos\beta_1\theta-1}{\beta_1^2}\frac{\cos\beta_2\phi-1}{\beta_2^2}\Bigg]\sin S(t-2\tau-\phi+\theta).
\end{aligned} \quad \text{(I-12)}$$

In this expression R_1 and R_2 are proportional to the rf fields due to the first and second pulses, respectively, and S is proportional to the field inhomogeneity. Various cases can be considered, depending on the relative magnitudes of these three quantities. In the usual nuclear induction experiments, for example, R_1 and $R_2\gg S_{\max}$. This leads to

$$\frac{w_0}{2}\sin R_1\theta(1-\cos R_2\phi)\cos S(t-2\tau-\phi+\theta), \quad \text{(I-13)}$$

which (neglecting ϕ and θ) is exactly the expression obtained by Hahn for the response near $T=2\tau$ and an inhomogeneity S. The total response may be obtained by integrating over all $S\ll R_1, R_2$. If a Gaussian inhomogeneity is assumed the result is again the same as Hahn's. Obviously, however, the limited Fourier frequency components of R_1 contained in S will not permit an accurate reproduction of the shape of R_1. We must consider the case where S is very large. If we take S large compared to both R_1 and R_2, we get many terms

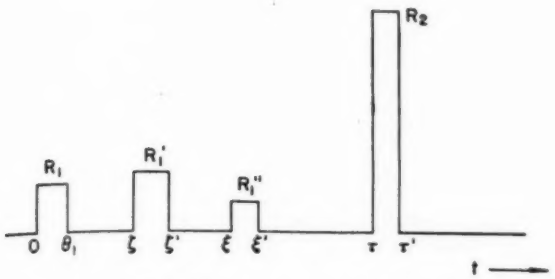


FIG. 12. Definition of complex code quantities.

containing S so that the final integral over S is not only complicated but shows a good deal of distortion of the image.

The most promising condition is that $R_2 \gg S \gg R_1$. This condition leads to

$$\frac{R_1 w_0}{2S}(1-\cos R_2\phi)\,\sin S\theta\,\cos S(t-2\tau-\phi+\theta)$$

$$-\frac{R_1 w_0}{2S}(1-\cos R_2\phi)(\cos S\theta-1)\,\sin S(t-2\tau-\phi+\theta)$$

$$=\frac{R_1 w_0}{2}(1-\cos R_2\phi)\left[\frac{\sin S(t-2\tau-\phi)}{S}-\frac{\sin S(t-2\tau-\phi+\theta)}{S}\right]. \quad \text{(I-14)}$$

If the integration is now performed over all S from $-\infty$ to $+\infty$, we get a square shaped image, the leading edge of which is given by the second term and occurs at $t=2\tau+\phi-\theta$, the trailing edge by the first term and occurs at $t=2\tau+\phi$. Thus we have the pattern shown in Fig. 11. The width of the image is exactly the same as that of the input pulse and has maximum amplitude for $R_2\phi=\pi$. Note also that the image distance measured from trailing edge of R_2 is the same as the object distance measured from the leading edge of R_2.

The conditions for good reproduction may be summarized briefly as

$$R_1 \ll S_{\max} \quad \text{or} \quad H_1 \ll \Delta H_{\max}$$

$$R_2\phi=\gamma H_2\phi \approx \pi,$$

and

$$S_{\max}\theta \gg 1.$$

In Fig. 3 we have plotted the echo response for various choices of $S_{\max}$, assuming the conditions above and $\theta=200$ μsec.

It is easy to make $R_1 \ll S_{\max}$, but it may require a great deal of power to make $R_2 \gg S_{\max}$. It is therefore advisable to see what happens if $R_2 \approx S_{\max}$. Rewriting the terms which give echoes near $t=2\tau$ with $R_1 \ll S_{\max}$, but no restriction on R_2, we obtain

$$\frac{R_1 w_0}{2S}\left[\frac{S^2}{\beta_2^2}(1-\cos\beta_2\phi)-(1-\cos\beta_2\phi)\right]$$

$$\times[\sin S(t-2\tau-\phi)-\sin S(t-2\tau-\phi+\theta)]. \quad \text{(I-15)}$$

The major effect of the correction term $S^2/\beta_2^2(1-\cos\beta_2\phi)$ is such as to reduce the amplitude of the echo if $S_{\max}\approx R_2$. If we choose $S_{\max}^2/R_2^2 \ll 1$, we can neglect this term, so that in practice it is sufficient to make $S_{\max}\approx R_2/3$.

Two other conditions must be satisfied if we are to reproduce coded signal, namely, that interpulse spacings be preserved and that no part of the code reproduce any other part. To show that this is actually so we consider the more general case of three small pulses irregularly separated, mirrored by a large fourth pulse. This is done in the same way as the previous calculation where now quantities are defined as in Fig. 12.

The resulting expression for $v(t)$ then is

$$v(t)=-w_0\exp\left(-\frac{t-\tau-\phi}{T}\right)\left[\sin R_2\phi\,\cos S(t-\tau-\phi)+\frac{S}{R_2}(\cos R_2\phi-1)\,\sin S(t-\tau-\phi)\right]$$

$$+\frac{R_1 w_0}{2S}\exp\left(-\frac{t-\theta_1-\theta_2-\theta_3}{T}\right)[-(\cos R_2\phi+1)\,\sin S(t-\phi)-(1-\cos R_2\phi)\,\sin S(t-2\tau-\phi)$$

$$+(1+\cos R_2\phi)\,\sin S(t-\theta_1-\phi)+(1+\cos R_2\phi)\,\sin S(t-2\tau+\theta_1-\phi)]$$

$$+\frac{R_1' w_0}{2S}\exp\left(-\frac{t-\zeta-\theta_2-\theta_3}{T}\right)[(1+\cos R_2\phi)\,\sin S(t-\zeta-\theta_2-\phi)+(1-\cos R_2\phi)\,\sin S(t-2\tau+\zeta+\theta_2-\phi)$$

$$-(1+\cos R_2\phi)\,\sin S(t-\zeta-\phi)-(1-\cos R_2\phi)\,\sin S(t-2\tau+\zeta-\phi)]$$

$$+\frac{R_1'' w_0}{2S}\exp\left(-\frac{t-\xi-\theta_3}{T}\right)[(1+\cos R_2\phi)\,\sin S(t-\xi-\theta_3-\phi)+(1-\cos R_2\phi)\,\sin S(t-2\tau+\xi+\theta_3-\phi)$$

$$-(1+\cos R_2\phi)\,\sin S(t-\xi-\phi)-(1-\cos R_2\phi)\,\sin S(t-2\tau+\xi-\phi)]. \quad \text{(I-16)}$$

If we consider only those terms containing 2τ and make $R_2\phi\approx\pi$, we see that we get echoes with amplitudes proportional to the input pulse amplitudes at the positions indicated in Fig. 13.

Here we have assumed the field inhomogeneity to extend from $-\infty$ to $+\infty$. Both pulse spacing and width are preserved. Also it is seen from (I-17) that under the conditions assumed for this calculation no undesirable echoes are produced.

A second procedure for storing information depends upon the so-called stimulated echo, reported by Hahn. In this case the information comes back in proper order and can be stored for times up to the relaxation time T_1. The capacity for storage, however, is limited by T_2.

The analysis is similar to that given in the preceding sections. In the following calculations we shall consider the case of two pulses of information, each being arbitrary but small rf signals. The information pulses must now be inserted after a large starting pulse, and as before are recalled by a strong reading pulse. The response is immediate, that is, the reading recalls the exact pattern that occurs after the starting pulse. Pictorially the situation is shown in Fig. 14.

Neglecting the decay terms in the Bloch equations for the intervals during pulses and considering R_2, $R_2' \gg S$; R_1, $R_1' \ll S$ we obtain, for $t \geqq \tau$,

$$
\begin{aligned}
v(t) = {} & \exp\left(-\frac{t-\theta_1-\theta_2-\phi_1-\phi_2}{T_2}\right) \\
& \frac{Sw_0}{2R_2}(\cos R_2\phi_1 - 1)(\cos R_2'\phi_2 - 1)\sin S(t-2\tau-\phi_2+\phi_1) \\
& -\frac{w_0}{2}\sin R_2\phi_1(\cos R_2'\phi_2 - 1)\cos S(t-2\tau-\phi_2+\phi_1) \\
& -\frac{Sw_0}{R_2'}\sin R_2\phi_1 \sin R_2'\phi_2 \sin S(t-2\tau-\phi_2+\phi_1) \\
& -\exp\left(-\frac{t-\zeta'-\theta_2-\phi_2}{T_2}\right)\frac{R_1w_0}{2S}(\cos R_2'\phi_2 - 1)[\sin S(t-2\tau+\zeta'-\phi_2) - \sin S(t-2\tau+\zeta-\phi_2)] \\
& -\exp\left(-\frac{t-\zeta'-\theta_2-\phi_1}{T_2} - \frac{\zeta-\phi_1}{T_1}\right)\frac{R_1w_0}{2S}(\cos R_2\phi - 1)(\cos R_2\phi_1 - 1)[\sin S(t-2\tau+\zeta'-\phi_2) - \sin S(t-2\tau+\zeta-\phi_2)] \\
& -\exp\left(-\frac{t-\xi'-\phi_2}{T_2} - \frac{\xi'-\phi_1}{T_1}\right)\frac{R_1'w_0}{2S}(\cos R_2\phi_1 - 1)(\cos R_2\phi - 1)[\sin S(t-2\tau+\xi'-\phi_2) - \sin S(t-2\tau+\xi'-\phi_2)] \\
& -\exp\left(-\frac{t-\xi'-\phi_2}{T_2}\right)\frac{R_1'w_0}{2S}(\cos R_2\phi - 1)[\sin S(t-2\tau+\xi'-\phi_2) - \sin S(t-2\tau+\xi'-\phi_2)] \\
& -\exp\left(-\frac{t-\tau'}{T_2} - \frac{\tau-\theta_1-\theta_2-\phi_1}{T_1}\right)w_0(\cos R_2\phi_1 - 1)\sin R_2'\phi_2 \cos S(t-\tau') \\
& -\exp\left(-\frac{t-\tau'}{T_2} - \frac{\tau-\theta_1-\theta_2-\phi_1}{T_1}\right)\frac{Sw_0}{R_2'}(\cos R_2\phi_1 - 1)(\cos R_2'\phi_2 - 1)\sin S(t-\tau') \\
& -\exp\left(-\frac{t-\tau'}{T_2}\right)w_0 \sin R_2'\phi_2 \cos S(t-\tau') \\
& -\exp\left(-\frac{t-\tau'}{T_2}\right)\frac{Sw_0}{R_2'}(\cos R_2'\phi_2 - 1)\sin S(t-\tau') \\
& +\exp\left(-\frac{t-\tau'+\zeta-\phi_1}{T_2} - \frac{\tau-\zeta'-\theta_2}{T_1}\right)\frac{R_1w_0}{S}\sin R_2\phi_1 \sin R_2'\phi_2[-\sin S(t-\tau'-\zeta'+\phi_1) + \sin S(t-\tau'-\zeta+\phi_1)] \\
& +\exp\left(-\frac{t-\tau'+\xi-\theta_1-\theta_1}{T_2} - \frac{-\tau\xi'}{T_1}\right)\frac{R_1'w_0}{S}\sin R_2\phi_1 \sin R_2'\phi_2[-\sin S(t-\tau-\xi'+\phi_1) + \sin S(t-\tau'-\xi+\phi_1)].
\end{aligned}
$$

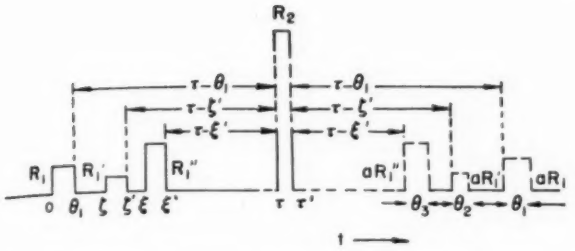


FIG. 13. Schematic diagram of echo response to complex code.

FIG. 14. Sequence of events when starting and reading pulses are used to recall code. Stimulated echoes are in proper sequence.

We eliminate the undesired echoes by making $R_2\phi_1 = R_2'\phi_2 = \pi/2$. Actually the information has to be inserted in a time short compared to T_1 to eliminate the undesired echoes completely. This condition is not as severe, however, as the one already imposed, that the information be inserted in a time short compared to T_2, since $T_2 < T_1$. The result then, when we keep highest-order terms at each echo, is

$$v(t)=\frac{w_0}{2}\exp\left(\frac{t-\theta_1-\theta_2-\phi_1-\phi_2}{T_2}\right)\cos S(t-2\tau-\phi_2+\phi_1)$$

$$-w_0\exp\left(-\frac{t-\tau'}{T_2}\right)\cos S(t-\tau')$$

$$+\frac{R_1 w_0}{S}\exp\left(-\frac{t-\tau'+\zeta-\phi_1}{T_2}-\frac{\tau-\zeta'-\theta_2}{T_1}\right)[\sin S(t-\tau'-\zeta+\phi_1)-\sin S(t-\tau'-\zeta'+\phi_1)]$$

$$+\frac{R_1' w_0}{S}\exp\left(-\frac{t-\tau'+\xi-\theta_1-\phi_1}{T_2}-\frac{\tau-\xi'}{T_1}\right)[\sin S(t-\tau'-\xi+\phi_1)-\sin S(t-\tau'-\xi'+\phi_1)]. \qquad \text{(I-17)}$$

The first term is responsible for the echo of the starting pulse at $t\cong 2\tau$, the second is the decay of the reading pulse, and the last two give us the echoes of the information pulses. It is obvious that the latter two echoes decay like $\exp(-\tau/T_1)$ and $\exp(-t/T_2)$, where $t=\zeta$ and ξ, respectively. Thus, if the information is inserted in time $t \ll T_2$, memory time is given by T_1.

Theory of Elasticity and Consolidation for a Porous Anisotropic Solid

M. A. Biot*
Shell Development Company, New York City, New York
(Received May 5, 1954)

The author's previous theory of elasticity and consolidation for isotropic materials [J. Appl. Phys. **12**, 155–164 (1941)] is extended to the general case of anisotropy. The method of derivation is also different and more direct. The particular cases of transverse isotropy and complete isotropy are discussed.

1. INTRODUCTION

THE theory of consolidation deals with the settlement under loading of a porous deformable solid containing a viscous fluid. In a previous publication[1] a consolidation theory was developed for isotropic materials. The purpose of the present paper is to extend the theory to the most general case of anisotropy. The method by which the theory is derived is also more general and direct. The same physical assumption is introduced, that the skeleton is purely elastic and contains a compressible viscous fluid. The theory may therefore also be considered as a generalization of the theory of elasticity to porous materials. It is applicable to the prediction of the time history of stress and strain in a porous solid in which fluid seepage occurs. The general equations derived in Sec. 2 are applied to the case of transverse isotropy in Sec. 3. This is a case of particular interest in the application of the theory to soils and natural rock formations, since transverse isotropic is the type of symmetry usually acquired by rock under the influence of gravity. For an isotropic material the equations reduce to a simple form given in Sec. 4. They are shown to coincide with the equations derived in reference 1. Application of the theory to specific cases was made previously,[2–4] and it was shown that the operational calculus offers a very powerful tool for the solution of consolidation problems in which a load is applied to the material at a given instant and the time history of the settlement is to be calculated. These methods are directly applicable to the more general nonisotropic case. More general solutions of the equations have been developed and will be presented in a forthcoming publication.

2. GENERAL EQUATIONS FOR THE ANISOTROPIC CASE

Let us consider an elastic skeleton with a statistical distribution of interconnected pores. This porosity is usually denoted by

$$f=V_p/V_b, \tag{2.1}$$

where V_p is the volume of the pores contained in a sample of bulk volume V_b. It is understood that the term "porosity" refers as is customary to the effective porosity, namely, that encompassing only the intercommunicating void spaces as opposed to those pores which are sealed off. In the following, the word "pore" will refer to the effective pores while the sealed pores will be considered as part of the solid. It will be noted that a property of the porosity f is that it represents also a ratio of areas

$$f=S_p/S_b, \tag{2.2}$$

i.e., the fraction S_p occupied by the pores in any cross-sectional area S_b of the bulk material. It must be assumed, of course, that the pores are randomly distributed in location but not necessarily in direction. That this relation holds may be ascertained by integrating S_p/S_b over a length of unity in a direction normal to the cross section S_b. The value of this integral then represents the fraction f of the volume occupied by the pores. It is seen that the ratio S_p/S_b is also independent of the direction of the cross section.

The stress tensor in the porous material is

$$\begin{Bmatrix} \sigma_{xx}+\sigma & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy}+\sigma & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz}+\sigma \end{Bmatrix}, \tag{2.3}$$

with the symmetry property $\sigma_{ij}=\sigma_{ji}$.

The partial components of this tensor do not have the conventional significance. If we consider a cube of unit size of the bulk material, σ represents the total normal tension force applied to the fluid part of the faces of the cube. Denoting by p the hydrostatic pressure of the fluid in the pores we may write

$$\sigma=-fp. \tag{2.4}$$

The remaining components σ_{xx}, σ_{xy}, etc., of the tensor are the forces applied to that portion of the cube faces occupied by the solid.

We shall now call our attention to this system of fluid and solid as a general elastic system with conservation properties. The solid skeleton is considered to have compressibility and shearing rigidity, and the fluid may be compressible. The deformation of a unit cube is assumed to be completely reversible. By deformation is meant here that determined by both strain tensors in the solid and the fluid which will now be defined. The average displacement components of the solid is designated by u_x, u_y, u_z, and that of the fluid by U_x, U_y, U_z.

* Consultant.
[1] M. A. Biot, J. Appl. Phys. **12**, 155–164 (1941).
[2] M. A. Biot, J. Appl. Phys. **12**, 426–430 (1941).
[3] M. A. Biot and F. M. Clingan, J. Appl. Phys. **12**, 578–581 (1941).
[4] M. A. Biot and F. M. Clingan, J. Appl. Phys. **13**, 35–40 (1942).

The strain components for the solid and the fluid, respectively, are

$$e_{xx}=\frac{\partial u_x}{\partial x} \quad \tfrac{1}{2}e_{xy}=\tfrac{1}{2}\left(\frac{\partial u_y}{\partial x}+\frac{\partial u_x}{\partial y}\right) \quad \text{etc.} \qquad (2.5)$$

$$\epsilon_{xx}=\frac{\partial U_x}{\partial x} \quad \tfrac{1}{2}\epsilon_{xy}=\tfrac{1}{2}\left(\frac{\partial U_y}{\partial x}+\frac{\partial U_x}{\partial y}\right) \quad \text{etc.}$$

By a generalization of the procedure followed in the classical theory of elasticity (5) we may write for the elastic potential energy V the expression

$$2V=\sigma_{xx}e_{xx}+\sigma_{yy}e_{yy}+\sigma_{zz}e_{zz}+\sigma_{yz}e_{yz}+\sigma_{zx}e_{zx}+\sigma_{xy}e_{xy}+\sigma\epsilon, \qquad (2.6)$$

with

$$\epsilon=\epsilon_{xx}+\epsilon_{yy}+\epsilon_{zz}.$$

If we assume that the seven stress components are linear functions of the seven strain components the expression $2V$ is a homogeneous quadratic function of the strain. This function is a positive definite form with twenty-eight distinct coefficients. The stress components are given by the partial derivatives of V as follows:

$$\partial V/\partial e_{xx}=\sigma_{xx} \quad \partial V/\partial e_{xy}=\sigma_{xy}, \quad \text{etc.}, \qquad (2.7)$$
$$\partial V/\partial\epsilon=\sigma.$$

This is written

$$\begin{pmatrix}\sigma_{xx}\\ \sigma_{yy}\\ \sigma_{zz}\\ \sigma_{yz}\\ \sigma_{zx}\\ \sigma_{xy}\\ \sigma\end{pmatrix}=\begin{pmatrix}c_{11}c_{12}c_{13}c_{14}c_{15}c_{16}c_{17}\\ c_{22}c_{23}c_{24}c_{25}c_{26}c_{27}\\ c_{32}c_{34}c_{35}c_{26}c_{37}\\ c_{44}c_{45}c_{46}c_{47}\\ c_{55}c_{56}c_{57}\\ c_{66}c_{67}\\ c_{77}\end{pmatrix}\begin{pmatrix}e_{xx}\\ e_{yy}\\ e_{zz}\\ e_{yz}\\ e_{zx}\\ e_{xy}\\ \epsilon\end{pmatrix}. \qquad (2.8)$$

Because the matrix of coefficients is that of a quadratic form we have the symmetry property

$$c_{ij}=c_{ji}. \qquad (2.9)$$

The total stress field (2.3) of the bulk material satisfies the equilibrium equations

$$\begin{aligned}&\frac{\partial}{\partial x}(\sigma_{xx}+\sigma)+\frac{\partial\sigma_{xy}}{\partial y}+\frac{\partial\sigma_{xz}}{\partial z}+\rho X=0;\\ &\frac{\partial\sigma_{yz}}{\partial x}+\frac{\partial}{\partial y}(\sigma_{yy}+\sigma)+\frac{\partial\sigma_{yz}}{\partial z}+\rho Y=0; \qquad (2.10)\\ &\frac{\partial\sigma_{zx}}{\partial x}+\frac{\partial\sigma_{zy}}{\partial y}+\frac{\partial}{\partial z}(\sigma_{zz}+\sigma)+\rho Z=0,\end{aligned}$$

where ρ is the mass density of the bulk material and X, Y, Z, the body force per unit mass. Substituting in (2.10) the stress components as functions of the strains from (2.8) we obtain three equations for the six unknown displacement $u_x\cdots U_x\cdots$. Three further equations between these unknowns are obtained by introducing the law governing the flow of a fluid in a porous material.

We introduce here a generalized form of Darcy's law for a nonisotropic material

$$\begin{pmatrix}-\partial p/\partial x+\rho_f X\\ -\partial p/\partial y+\rho_f Y\\ -\partial p/\partial z+\rho_f Z\end{pmatrix}=\begin{pmatrix}k_{xx} & k_{xy} & k_{xz}\\ k_{yx} & k_{yy} & k_{yz}\\ k_{zx} & k_{zy} & k_{zz}\end{pmatrix}\begin{pmatrix}\dot U_x-\dot u_x\\ \dot U_y-\dot u_y\\ \dot U_z-\dot u_z\end{pmatrix}, \qquad (2.11)$$

where ρ_f is the mass density of the fluid. The matrix k_{ij} constitutes a generalization of Darcy's constant if we include in it the viscosity coefficient. The average velocities of the fluid and solid are denoted by $\dot U_x\cdots\dot u_x\cdots$.

The symmetry of the coefficients

$$k_{ij}=k_{ji} \qquad (2.12)$$

results from the existence of a dissipation function such that the rate of dissipation of the energy in the porous material at rest is expressed by the positive definite quadratic form

$$2D=\sum^{ij}k_{ij}\dot U_i\dot U_j. \qquad (2.13)$$

If we multiply Eq. (2.11) by f and take (2.4) into account we obtain

$$\begin{pmatrix}\partial\sigma/\partial x+\rho_1 X\\ \partial\sigma/\partial y+\rho_1 Y\\ \partial\sigma/\partial z+\rho_1 Z\end{pmatrix}=\begin{pmatrix}b_{xx} & b_{xy} & b_{xz}\\ b_{yx} & b_{yy} & b_{yz}\\ b_{zx} & b_{zy} & b_{zz}\end{pmatrix}\begin{pmatrix}\dot U_x-\dot u_x\\ \dot U_y-\dot u_y\\ \dot U_z-\dot u_z\end{pmatrix}, \qquad (2.14)$$

with $\rho_1=\rho_f f=$ the mass of fluid per unit volume of bulk material. The three equations obtained by combining (2.10) and (2.8) in addition to the three Eqs. (2.14) determine the six unknown displacement components for the fluid and the solid.

3. THE CASE OF TRANSVERSE ISOTROPY

The above equations are valid for the most general case of a symmetry. In practice, however, materials will be either isotropic or exhibit a high degree of symmetry which greatly simplifies the equations. Let us consider first the case of a material which is axially symmetric about the z axis. This type of symmetry is referred to by Love[6] as transverse isotropy (page 160). The expression for the strain energy in this case is

$$\begin{aligned}2V=&(A+2N)(e_{xx}+e_{yy})^2+Ce_{zz}^2+2F(e_{yy}+e_{xx})e_{zz}\\ &+L(e_{yz}^2+e_{zx}^2)+N(e_{xy}^2-4e_{xx}e_{yy})\\ &+2M(e_{xx}+e_{yy})\epsilon+2Qe_{zz}\epsilon+R\epsilon^2. \qquad (3.1)\end{aligned}$$

This expression is invariant under a rotation around the z axis. It is written in such a way as to bring out expressions such as $e_{xy}^2-4e_{xx}e_{yy}$ and $e_{xx}+e_{yy}$ which are invariant under a rotation about the z axis. The coefficient $A+2N$ is written this way for reasons of conformity.

[6] A. E. H. Love, *A Treatise On the Mathematical Theory of Elasticity* (Dover Publications, New York, 1944).

Since A does not appear in any other term, the quantity $A+2N$ is an independent coefficient which could have been written as P [see (4.5)]. The stress-strain relations derived from (2.7) and (3.1) are

$$\begin{aligned}
\sigma_{xx}&=2Ne_{xx}+A(e_{xx}+e_{yy})+Fe_{zz}+M\epsilon;\\
\sigma_{yy}&=2Ne_{yy}+A(e_{xx}+e_{yy})+Fe_{zz}+M\epsilon;\\
\sigma_{zz}&=Ce_{zz}+F(e_{xx}+e_{yy})+Q\epsilon;\\
\sigma_{yz}&=Le_{yz};\\
\sigma_{zx}&=Le_{zx};\\
\sigma_{xy}&=Ne_{xy};\\
\sigma&=M(e_{xx}+e_{yy})+Qe_{zz}+R\epsilon.
\end{aligned}\tag{3.2}$$

There are therefore in this case eight elastic coefficients.

The equations of flow contain two coefficients of permeability, one in the z direction, the other in the x, y plane, and may be written

$$\begin{aligned}
\partial\sigma/\partial x+\rho_1X&=b_{xx}(\dot U_x-\dot u_x);\\
\partial\sigma/\partial y+\rho_1Y&=b_{xx}(\dot U_y-\dot u_y);\\
\partial\sigma/\partial z+\rho_1Z&=b_{zz}(\dot U_z-\dot u_z).
\end{aligned}\tag{3.3}$$

These equations along with the stress-strain relation (3.2) and the equilibrium relations (2.10) yield six equations for the six displacement components in the case of transverse isotropy.

4. THE CASE OF ISOTROPY

In the case of complete isotropy the strain energy function (3.1) becomes

$$\begin{aligned}
2V=&(A+2N)(e_{xx}+e_{yy}+e_{zz})^2\\
&+N(e_{yz}^2+e_{zx}^2+e_{xy}^2-4e_{yy}e_{zz}\\
&-4e_{zz}e_{xx}-4e_{xx}e_{yy})\\
&+2Q(e_{xx}+e_{yy}+e_{zz})\epsilon+R\epsilon^2.
\end{aligned}\tag{4.1}$$

We put

$$e=e_{xx}+e_{yy}+e_{zz}.\tag{4.2}$$

The stress-strain relations derived from (2.7) are

$$\begin{aligned}
\sigma_{xx}&=2Ne_{xx}+Ae+Q\epsilon;\\
\sigma_{yy}&=2Ne_{yy}+Ae+Q\epsilon;\\
\sigma_{zz}&=2Ne_{zz}+Ae+Q\epsilon;\\
\sigma_{yz}&=Ne_{yz};\\
\sigma_{zx}&=Ne_{zx};\\
\sigma_{xy}&=Ne_{xy};\\
\sigma&=Qe+R\epsilon.
\end{aligned}\tag{4.3}$$

There are in this case four elastic constants, and this checks with the result obtained in reference 1. The equations of flow contain a single coefficient b. They are written

$$\begin{aligned}
\partial\sigma/\partial x+\rho_1X&=b(\dot U_x-\dot u_x);\\
\partial\sigma/\partial y+\rho_1Y&=b(\dot U_y-\dot u_y);\\
\partial\sigma/\partial z+\rho_1Z&=b(\dot U_z-\dot u_z).
\end{aligned}\tag{4.4}$$

We shall assume that there is no body force and put $X=Y=Z=0$. Substitution of expression (4.3) into the equilibrium Eq. (2.10) for the stresses and the flow Eq. (4.4) yield the six equations

$$\begin{aligned}
&N\nabla^2\bar u+(P-N+Q)\ \mathrm{grad}e+(Q+R)\ \mathrm{grad}\epsilon=0\\
&\mathrm{grad}(Qe+R\epsilon)=b(\partial/\partial t)(\bar U-\bar u).
\end{aligned}\tag{4.5}$$

We have put $P=A+2N$.

Taking the divergence of the second equation we may also write

$$\begin{aligned}
&N\nabla^2\bar u+(P-N+Q)\ \mathrm{grad}e+(Q+R)\ \mathrm{grad}\epsilon=0\\
&Q\nabla^2e+R\nabla^2\epsilon=b(\partial/\partial t)(\epsilon-e).
\end{aligned}\tag{4.6}$$

In the previous theory (1) we had obtained these equations by a different method and in a different form. To show their equivalence we write the stress-strain relations by eliminating ϵ from Eqs. (4.3)

$$\begin{aligned}
\sigma_{xx}&=2Ne_{xx}+\left(A-\frac{Q^2}{R}\right)e+\frac{Q}{R}\sigma;\\
\sigma_{yy}&=2Ne_{yy}+\left(A-\frac{Q^2}{R}\right)e+\frac{Q}{R}\sigma;\\
\sigma_{zz}&=2Ne_{zz}+\left(A-\frac{Q^2}{R}\right)e+\frac{Q}{R}\sigma;\\
\sigma_{yz}&=Ne_{yz};\\
\sigma_{zx}&=Ne_{zx};\\
\sigma_{xy}&=Ne_{xy}.
\end{aligned}\tag{4.7}$$

Substituting these in the equilibrium relation (2.10) we find

$$\begin{aligned}
&N\nabla^2\bar u+[P-N-Q^2/R]\ \mathrm{grad}e\\
&\qquad+(Q+R)/R\ \mathrm{grad}\sigma=0.
\end{aligned}\tag{4.8}$$

We also derive from (4.4)

$$\nabla^2\sigma=b\frac{\partial}{\partial t}(\epsilon-e)=\frac{b}{R}\frac{\partial\sigma}{\partial t}-b\frac{Q+R}{R}\frac{\partial e}{\partial t}.\tag{4.9}$$

Equations (4.8) and (4.9) are in the form obtained in reference 1. We note that the significance of σ in that reference is equivalent to $-\sigma/f$ in our present notation.

Consider now the case of an incompressible material. This corresponds to the condition

$$e(1-f)+f\epsilon=0.\tag{4.10}$$

Since this must be satisfied for all values of σ we derive from the last relation (4.3) that both R and Q are infinite with the condition

$$Q/R=(1-f)/f.\tag{4.11}$$

Since $A-Q^2/R=S$ must remain finite the stress strain

law becomes

$$\sigma_{xx}=2Ne_{xx}+Se+\frac{1-f}{f}\sigma;$$

$$\sigma_{yy}=2Ne_{yy}+Se+\frac{1-f}{f}\sigma;$$

$$\sigma_{zz}=2Ne_{zz}+Se+\frac{1-f}{f}\sigma; \qquad (4.12)$$

$$\sigma_{yz}=Ne_{yz};$$

$$\sigma_{zx}=Ne_{zx};$$

$$\sigma_{xy}=Ne_{xy}.$$

Substituting these expressions in the equilibrium relations (2.10) we derive

$$N\nabla^2\bar{u}+(N+S)\,\mathrm{grad}e+(1/f)\,\mathrm{grad}\sigma=0 \qquad (4.13)$$

and from (4.9)

$$\nabla^2\sigma=-\frac{b}{f}\frac{\partial e}{\partial t}. \qquad (4.14)$$

Taking the divergence of (4.13)

$$(2N+S)\nabla^2e+\frac{1}{f}\nabla^2\sigma=0. \qquad (4.15)$$

Hence (4.14) may be written

$$f^2(2N+S)\nabla^2e=b\frac{\partial e}{\partial t}. \qquad (4.16)$$

This is the equation of heat conduction. Equations (4.13) and (4.16) coincide with those obtained in reference 1 for the incompressible case.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 2 FEBRUARY, 1955

Radiation of Plasma Noise from Arc Discharge

T. Takakura, K. Baba, K. Nunogaki, and H. Mitani
Osaka City University, Kitaku, Osaka, Japan

(Received June 8, 1954)*

The mechanism of radiation of plasma noise inherent to arc discharge is studied experimentally. Intense electromagnetic radiation of random noise character is observed at the frequency of plasma oscillation of the ion sheath formed at the cathode drop. Maximum radiation occurs when the external circuit connected to the cathode is in resonance with the plasma frequency. The oscillating current in the external circuit is concluded to be generated by the periodic electron emission from the cathode caused by small perturbation of potential gradient at the cathode surface as a result of the variation in the ion sheath potential. This conclusion is supported by several experiments under various air pressures, electrode materials, current densities, etc.

1. INTRODUCTION

COLD cathode dc arc discharge in air at ordinary atmospheric pressure was found to radiate intense radio noise in the microwave region equivalent to thermal radiation of 10^6 to 10^7 °K at 0.5 to 5 amperes of discharge current without any external resonance circuit. The radiation was observed at frequencies of 3300 mc, 190 mc, 15 mc, and 1.5 mc under various air pressures with various electrode materials and shapes by detecting the noise involved therein.

More or less similar investigations were conducted by Melloh and others.[1] However, little seems to have been done in the explanation of the mechanism of noise radiation. The present paper deals with the explanation of the generation of such noise radiation.

* The original unrevised manuscript was received September 8, 1953.

[1] A. M. Melloh, Proc. Inst. Radio Engrs. **28**, 179 (1940); I. Matsumoto, Ôyô-Butsuri **15**, 21 (1946); I. Hayashi and E. Abe, Rikoken-Hôkoku **3**, 9 (1949); Davies, Pear, and White, Electronics **23**, 96 (1950).

2. EXPERIMENTAL RESULTS

(i) The Noise of Arc Discharge at an Atmospheric Pressure

Radiation of noise from dc arc discharge in air was observed by using a 3300 mc receiver with a band width of approximately 4 mc connecting to a horn antenna. The condition of discharge was changed by the shape and material of the electrodes and the discharge current. The radio noise radiated from the arc discharge was received by the horn antenna as shown in Fig. 1(a).

(a) *Influence of Electrode Material*

Relative intensity of radio noise at 3300 mc from the dc arc discharge at 1 ampere with electrodes of various materials was measured and tabulated in Table I. Care was taken to make each electrode of different materials in a similar shape. As seen in Table I, the positive electrode was found to be of little significance in determining the intensity of the noise. When a carbon or mercury electrode was used as the cathode, an excep-

TABLE I. Relative intensity of 3300 mc radio noise from arc discharge in air with electrodes of various materials.

Negative electrode \ Positive electrode	Cu	Fe	Al	Mo	W	Pb	Hg	C
Cu	4	4	3	3	3	3	3	3
Fe	13	13	15~11	13~8	15~7	10~7	13	13
Al	33	33	35	35	35~30	25~20	30	30
Mo	11	8	7	11	10	9	10	9
W	10	7	10	7	10	6	7	8
Pb	<0.1	<0.1	...	...	...	...	...	...
Hg	<0.05	<0.05	...	...	...	...	...	...
C	<0.05	<0.05						

tionally low noise equivalent to a thermal emission of not more than 10^4 °K was observed.

(b) *Influence of External Circuit*

Investigation was made as to whether the generation and radiation of noise were due essentially to the external circuit or not. Though the radiation of the noise from the electrodes could not be eliminated completely, it could be minimized by the use of electrodes as shown in Fig. 1(a). The absolute intensity of noise was measured at 3300 mc by comparing with a standard noise generator, i.e., with the thermal radiation of 10^4 °K from a fluorescent lamp.

A 2-mm copper gap with a discharging current of 1 ampere dc radiated noise which was equivalent to approximately 10^6 °K when the electric polarity of the receiving horn was parallel to the discharging current, and 10^5 °K when it was perpendicular. This indicates that there is polarization of the noise field along the discharging current.

Electrodes with half wavelength dipole as illustrated in Fig. 1(b) radiated noise whose intensity was equivalent to approximately 10^7 °K and 10^5 °K for parallel and perpendicular polarization, respectively, as in the preceding case.

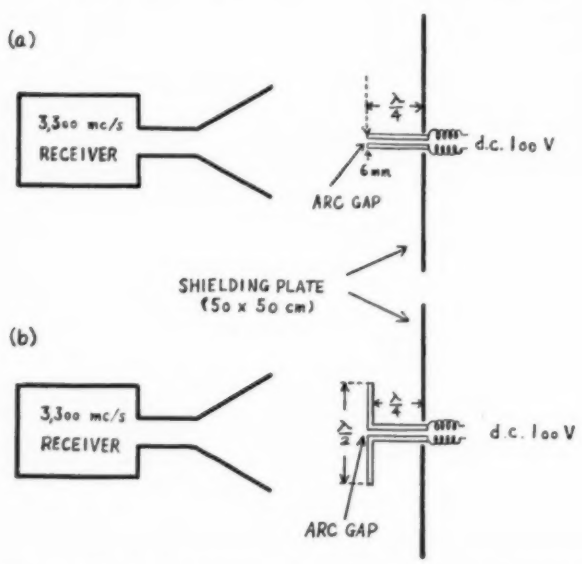


FIG. 1. Experimental layout of the arc noise intensity measurement. (*a*) The radiation from electrode is minimum. (*b*) The radiation from electrodes is maximum.

The noise intensity increased by increasing the arc gap from 0.5 mm to 3 mm in the case of Fig. 1(a), but it did not change in the case of Fig. 1(b). Field distribution along the dipole was measured by a probe, as shown in Fig. 2(a). This was done at 3300 mc, 190 mc, and 15 mc at discharging current of dc 2 amperes and 50 amperes. Symmetrical distribution of noise voltage on both electrodes, as shown in Fig. 2(b), was observed at 3300 mc when the arc gap was shorter than 10 mm, but at gaps greater than 10 mm, the voltage could be detected only along the cathode electrode (Fig. 2(c)). However, symmetrical voltage distribution on both electrodes was observed at lower frequencies even at such increased gaps. No remarkable change in intensity was observed at all frequencies when the arc gap was increased gradually from 0.5 mm to 15 mm.

(c) *Influence of Discharge Current and the Diameter of the Electrodes*

The intensity of noise at 3300 mc was measured by varying the discharging current and the diameter of electrodes made of copper and iron. As had been expected, the diameter of the anode had no effect on the noise intensity; whereas the effect of cathode diameter was observed to exist. This is shown in Table II. Under the particular condition shown in Table II, the weaker the arc current and the larger the diameter of the cathode electrode, the stronger was the 3300 mc noise.

(d) *Influence of Magnetic Field*

The polarity and the intensity of the noise were seldom affected by superposing a magnetic field of 10^3

TABLE II. Relative intensity of 3300 mc radio noise from arc discharges in air with various electrode-diameter and discharge currents.

Cathode diameter \ Discharge current	Copper arc 1.5 A	0.7 A	Iron arc 1.5 A	1.1 A	0.9 A
2.5 mm	25	40	15	17	30
1.5	17	25	7	10	21
1.0	6	14	...	...	...

gauss along the discharging current. The gyro-magnetic flux density B (gyro-frequency $f_\Omega = eB/2\pi m$ in MKS units) at 3000 mc for electron and copper ion is 10^3 and 10^8 gauss, respectively.

(ii) The Noise of Arc Discharge in Air at Reduced Pressure

The variation of relative noise intensity *versus* decreasing air pressure with various cathode material was observed at 3300 mc, (band width 4 mc), 190 mc (band width 4 mc), 15 mc (band width 2 mc), and 1.5 mc (band width 10 kc). Antennas for all the frequencies, except for 3300 mc, were made of pieces of wires attached to the end of flexible cables. These were placed near the discharge tube. The arc gap was fixed at 0.5 mm to 1 mm. Discharge was started by a dc high voltage. Discharge current of about 0.5 ampere dc was drawn for a few seconds to avoid the change of

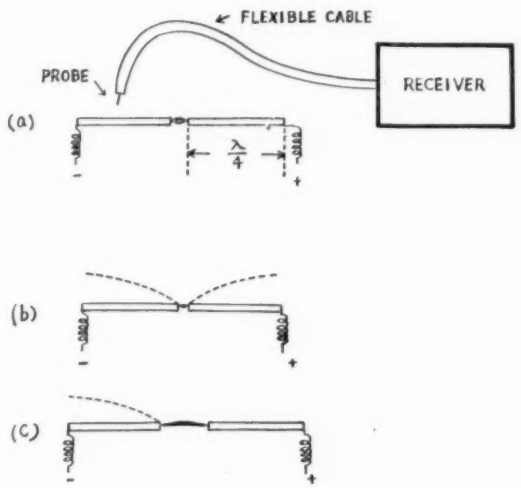


FIG. 2. Measurement of voltage distribution of arc discharge noise on the electrodes. (*a*) The voltage distribution was measured by the probes of coaxial cables at 3300 mc, 190 mc, and 15 mc. (*b*) The voltage distribution of the noise (dotted line), when the arc gap was shorter than 10 mm at 3300 mc. (*c*) The voltage distribution, when the arc gap was longer than 10 mm at 3300 mc.

noise intensity because of the rise in temperature and oxidation of the cathode.

The result of observation is shown in Fig. 3, where the peaks of noise are seen to shift toward lower frequency with the decrease of air pressure. The noise frequency is lower in the case of the carbon and mercury arc than that of the other metallic arc at the same pressure. Since current density is known to decrease with decreasing pressure, and current density of the carbon and mercury arc is also known to be lower than that of other metallic cathodes,[2] it can be said that within the range of the observed condition the lower the current density is, the lower will be the frequency of noise.

When a mercury pool was used as a cathode, excess

[2] J. D. Cobine, *Gaseous Conductors* (McGraw-Hill Book Company, Inc., New York, 1941), first edition, p. 301.

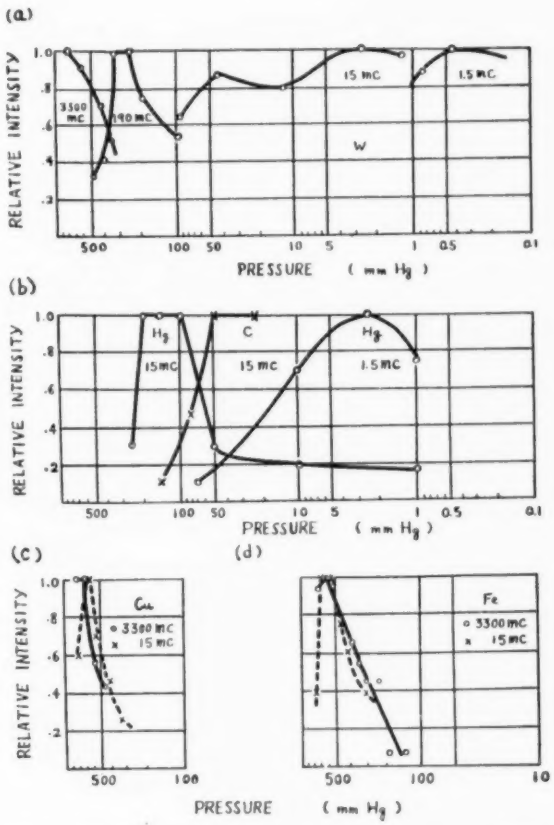


FIG. 3. Noise intensity *versus* air pressure for various cathode materials and frequencies. (The anode is iron.) (*a*) Tungsten arc discharge with 2-mm diameter cathode and 1-mm arc gap at 0.4 A dc. (*b*) Mercury and carbon arc, at 0.3 A dc. (*c*) Copper arc with 4-mm diameter cathode and 1-mm gap at 0.4 A dc. (*d*) Iron arc with 2.5-mm diameter cathode and 1-mm gap at 0.5 A dc.

noise was not observed at 3300 mc and 190 mc under the discharging condition shown in Fig. 3(b). However, when the tube was evacuated and the temperature of the mercury pool rose to 70°C after a discharge current of 1 to 2 amperes was drawn continuously, the 3300 mc noise was observed together with a weak 190 mc noise. This noise may be attributed to the plasma "ion" oscillation at the cathode drop, which will be discussed in the next section. In the same manner, the plasma "electron" oscillation is likely to generate the 3300 mc noise at a lower pressure. Glow discharge in the usual Geissler tube, 3 cm in diameter with copper electrodes separated 10 cm apart, radiated the 3300 mc noise together with the 1.5 mc noise at a pressure of the order of 10^{-1} mm Hg at a discharging current of dc 1000 milliamperes. The location of the noise source was investigated by sliding a small horn antenna of the 3300 mc receiver along the discharge tube. Radiation of noise took place only near the cathode. The intensity of the 3300 mc noise was of the order of 10^6 °K. The polarity was not clear. The cause of the 3300 mc and 1.5 mc noise may be attributed to the "electron" and "ion" oscillation at the cathode drop region, respectively.

3. DISCUSSION

From the foregoing experimental evidence, the following hypothesis is offered. (1) The noise source is near the cathode. (2) The noise is due to plasma oscillation. To proceed theoretically with this hypothesis, one needs to make an evaluation of the electron and ion density at the cathode drop which is difficult to measure because the temperature is high and the cathode drop region is very small. Therefore, estimation of charge density will be made by the well-known current density value[2,3] at the cathode spot. The estimated electron and ion density of tungsten arc discharge at 1 atmos and mercury arc at high vacuum are of the order of 10^{15} electrons per cubic centimeter and 10^{16} ions per cubic centimeter, respectively. The plasma density n (plasma frequency $f_p=1/2\pi(ne^2/m\epsilon_0)^{\frac{1}{2}}$, in MKS units) of electron and tungsten ion or mercury ion at 3000 mc is of the order of 10^{11} electrons per cubic centimeter and 10^{16} ions per cubic centimeter, respectively. With this evidence and reference, the origin of noise may be attributed to the plasma ion oscillation at the cathode drop region, which in turn will reasonably account for the experimental result of Sec. 2(d). It must be recalled at this point that the so-called plasma oscillation is usually a longitudinal oscillation which does not radiate directly any transversal wave by itself. Therefore, the radiation of noise is probably the result of the periodic change of electron emission current from the cathode spot controlled by the fluctuation of potential caused by the plasma oscillation of ion sheath at the cathode spot region. In other words, the emission current from the cathode changes slightly in a pulsive manner by the periodical variation of the cathode field strength. Only an oscillating perturbation of 10^{-6} part of the cathode field strength is necessary to account for the intensity of the noise measured at 3300 mc (see Appendix). Here, the plasma oscillation is understood to be maintained by the excitation of the well-defined electron beam emitted from the cathode as treated by several authors.[4–7]

According to J. D. Cobine and C. J. Gallagher,[8] the generation of lower frequency noise in glow discharge tubes is caused by the velocity modulation of the electron beam by the plasma oscillation, whereby the electron beam subject to the ion plasma oscillation is bunched along its passage to the anode. If this were the actual case, one should have observed an intense oscillation at the anode rather than at the cathode. Also, periodic change in oscillation intensity should have resulted by charging the arc gap. As is seen from Sec. 2(b), the actual result is quite contrary to this.

One of the results of Sec. 2(b), stating that the noise intensity increased with increasing arc gap in the case of Fig. 1(a), may be accounted as follows. As is mentioned in the Appendix, the electrodes play an important role in the "intensity" of radiation when the thickness of the cathode drop is small compared with the length of the electrodes. Under this condition, to increase the arc gap is to increase the dipole moment, the situation corresponding to Fig. 1(a). Whereas, in the case of Fig. 1(b), in which a $\frac{1}{2}\lambda$ dipole exists, it is reasonable that a small change in the arc gap has little influence on the radiation intensity.

According to Fig. 3, peaks of noise for each of the observed frequencies appeared separately along the pressure axis in the case of tungsten, carbon, and mercury cathodes. Remarkable difference is seen for copper and iron cathodes, where peaks of noise for different observed frequencies fell on the same curve Fig. 3(c) and (d).† This situation may occur because tungsten and mercury ions are extremely heavy and carbon ion is approximately equal in mass compared with oxygen or nitrogen ion mixed with them, resulting in a better defined group of ion density than that of copper or iron ions whose ion mass lies intermediate to tungsten and carbon.

The 3300 mc noise radiated from the Geissler tubes mentioned in the last part of Sec. (ii) may be attributed to the "electron" plasma noise. It may be said that not only plasma "ion" oscillation but plasma "electron" oscillation also has the same effect in the generation and radiation of noise as expected by the mechanism of radiation mentioned above.

4. CONCLUSIONS

As a result of the experimental observation and theoretical insight into the phenomena of electromagnetic radiation of noise from arc discharge, the following items may be concluded.

(1) The frequency of noise radiation shifts toward lower frequency by decreasing the pressure of air in which the arc discharge takes place. (2) The frequency of radiation is determined by the current density at the cathode drop. The ion plasma frequency of the ion sheath at the cathode drop is of the same order of the noise radiation. (3) The source of the noise is near to the cathode. The anode has little to do in the generation of noise. (4) Generation of noise radiation is probably caused by the periodic change of electron emission from the cathode spots controlled by the plasma ion oscillation of the ion sheath at the cathode drop. (5) Radiation of noise wave is emphasized by the presence of electrodes, especially by the cathode electrode when the arc gap is large. (6) A small perturbation of cathode field strength of the order of 10^{-4} percent is sufficient to account for the measured noise intensity of dc copper arc discharge in air.

[3] J. D. Cobine and C. J. Gallagher, Phys. Rev. **74**, 1524 (1948).
[4] D. Bohm and E. P. Gross, Phys. Rev. **75**, 1864 (1949).
[5] D. Bohm and E. P. Gross, Phys. Rev. **79**, 922 (1950).
[6] J. Feinstein and H. K. Sen, Phys. Rev. **83**, 405 (1951).
[7] M. Sumi, J. Phys. Soc. Japan **9**, 88 (1954).
[8] J. D. Cobine and C. J. Gallagher, J. Appl. Phys. **18**, 110 (1949).

† In order to avoid confusion, curves at 190 mc and 1.5 mc are not shown on the same graphs.

5. ACKNOWLEDGMENTS

We would like to express our thanks to Professor Yuzuru Watase for his encouragement and criticism, and particularly to Professor Sanai Mito for his helpful criticism during the preparation of this paper.

APPENDIX

Calculation of the intensity of radiation of radio waves resulting from the change of electron current controlled by the plasma oscillation of ion sheath at the cathode drop will be dealt with in the following.

The electric field strength Ec at the cathode in the case of arc discharge, when the space charge in the cathode-drop region is considered entirely the result of the positive ion, is given by[9]

$$E_c=\frac{4V_c}{3d}, \tag{1}$$

where V_c is the cathode drop and d is the thickness of the cathode drop region. The emission current density from the cathode is given by[10]

$$j=j_0\exp(4.4Ec^{\frac{1}{2}}/T), \tag{2}$$

where E_c is in volts/cm and j_0 is the thermionic emission for zero field strength at the cathode.

If E_c varies by a small amount $E_c(1+\xi)$, inherent variation in the emission current follows as

$$j=j_0\exp\{4.4E_c^{\frac{1}{2}}(1+\xi)^{\frac{1}{2}}/T\} \tag{3}$$

or

$$j\simeq j_0(1+\zeta)\exp(4.4E_c^{\frac{1}{2}}/T),$$

where $\zeta=(2.2E^{\frac{1}{2}}/T)\xi$. Substituting probable values for E_c and T evaluated as,[11] $E_c=10^6$ volts/cm, and $T=2500$ °K, yields

$$\zeta\simeq 0.9\xi. \tag{4}$$

This small change of E_c may either be caused by the change of V_c or d, whether separately or simultaneously, in Eq. (1) under the influence of the plasma oscillation of the ion sheath at the cathode drop. This pulsating current results in an electric dipole which contributes in the radiation of electric waves.

The radiation from a dipole with a dipole moment P_0 is given by[12]

$$W=\mu_0\omega^4 p_0^2/12\pi c \text{ watts}, \tag{5}$$

where $P_0=I_0\Delta l/\omega$. I_0, Δl, and ω are the amplitudes of ac current, the length of the dipole and the angular frequency, respectively. Equation (5) reduces to

$$W=\frac{\pi c\mu_0}{3}\left(\frac{\Delta l}{\lambda}\right)^2 I_0^2 \text{ watts}. \tag{6}$$

In the case of arc noise, it is assumed that electric current varies as $I_0e^{j\omega t}$ by the plasma oscillation of the ion sheath at the cathode drop region. The measured equivalent temperature of the noise source of arc discharge in question is related by

$$KT\Delta f=W=\frac{\pi c\mu_0}{3}\left(\frac{\Delta l}{\lambda}\right)^2 I_0^2, \tag{7}$$

where K is the Boltzmann constant and Δf is the band width of the receiver.

Taking $\Delta l/l=6.7\times10^{-2}$, $T=10^6$ °K at 3300 mc [in the case of Fig. 1(a)] and $\Delta f=4$ mc/sec one obtains the effective value of ac current

$$I_e=I_0/\sqrt{2}=4\times10^{-6}\text{A}. \tag{8}$$

Here, we have taken $\Delta l=6$ mm, i.e., the span between the outer side of the electrodes [see Fig. 1(a)]. Since the cathode-drop thickness is very thin the radiation from the end of the electrodes cannot be neglected, if the noise is caused by the change of electric current as we have considered. In this case, dc discharge current is 1 A, so the perturbation ζ of current is 4×10^{-6}. From Eq. (4) the required perturbation of E_c is also of the order of 10^{-6}.

The radiation from a $\lambda/2$ dipole antenna is given by

$$W=73I_e^2 \text{ watts}.$$

In the case of Fig. 1(b), setting $T=10^7$ °K at 3300 mc, we obtain

$$I_e=3\times10^{-6}\text{ A}.$$

This value is nearly equal to the above case (i.e. Eq. (8).)

The small perturbation, which is of the order of 10^{-6} of the cathode electric field strength, is sufficient to account for the measured noise intensity of arc discharge in air.

The radiation from electrodes cannot be neglected even if the electrodes are small, because the cathode-drop thickness is very thin in the usual case.

[9] See reference 2, first edition, p. 303.
[10] See reference 9, p. 118.
[11] See reference 9, pp. 303–305.
[12] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), first edition, p. 437.

Measurement of Carrier Lifetimes in Germanium and Silicon*

DONALD T. STEVENSON AND ROBERT J. KEYES
Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received April 2, 1954)

The decay of photoconductivity has been used to measure the lifetime of excess carriers in rectangular samples of germanium and silicon. The sample is illuminated by a short pulse of light and the sample lifetime obtained from an oscilloscope display of the decay of photoconductivity. Analysis of the solution of the diffusion equation yields methods of measuring the bulk lifetime, the surface recombination velocity, and the diffusion constant.

1. INTRODUCTION

THE lifetime of added current carriers is an important physical property of semiconductors such as germanium and silicon. It is of interest in studies of the basic energy level structure of semiconductors and is an important parameter in transistor technology.

Numerous methods for measuring carrier lifetime have been described.[1] In this paper we shall describe a method developed in this laboratory in which one observes the decay of photoconductivity in a semiconductor sample of known geometry. This method was also developed at the Bell Telephone Laboratories and is described briefly by Haynes and Hornbeck.[2] The method described by Navon, Bray, and Fan differs from the one here discussed only in that they inject the excess carriers electrically from leads placed on the sample rather than optically.

The experimental arrangement is shown in Fig. 1. A short pulse of light creates extra hole-electron pairs in the sample. The excess conductance of the sample under proper experimental conditions[3] is directly proportional to the total number of excess carriers present. For a small conductance change the voltage change at the terminals of the constant current source is proportional to the conductance change. This voltage, therefore, decays back to its steady state value with the same time dependence as the number of excess carriers in the germanium sample. The time dependence of the excess carrier density is obtained from the display of the voltage change on an oscilloscope. Analysis of the oscilloscope trace then yields the sample lifetime.

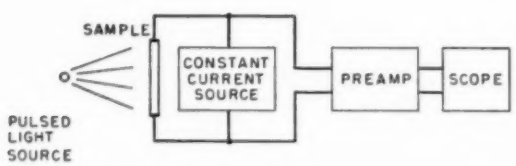


FIG. 1. Experimental arrangement.

2. APPARATUS

Two methods of obtaining the necessary short light pulses have been used. In the first, shown schematically in Fig. 2a, a 500-watt slide projector throws a beam of light onto a rotating metal mirror driven by a variable speed motor. The sample, placed a distance d from the mirror, is then illuminated for only a short time out of each cycle. A lens is sometimes used to focus the beam on the sample if greater intensity is needed.

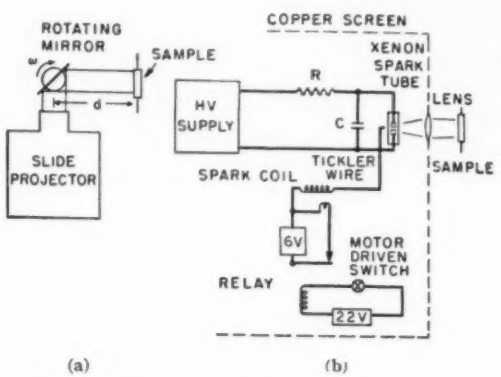


FIG. 2. Light sources: (a) rotating mirror source, (b) xenon spark source.

The pulse length in seconds is given by $\Delta t = (W+E)/2\omega d$, where ω is the angular velocity of the mirror, W is the width of the beam, and E is the sample dimension along the direction of sweeping or the lens aperture if a lens is used. With our rotating mirror setup the lowest lifetime measurable is about 10 μsec.

For shorter lifetimes a high pressure xenon spark tube is used as a light source.[4] A schematic diagram of the tube and its associated circuitry is shown in Fig. 2b. Condenser C is charged through resistance R from the high voltage supply (2000–4000 volts). A very high voltage pulse from a model airplane spark coil is put onto a small wire wrapped around the spark tube. This

* The research in this document was supported by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology, Cambridge, Massachusetts.

[1] Navon, Bray, and Fan, Proc. Inst. Radio Engrs. **40**, 1342 (1952); J. R. Haynes and W. Shockley, Phys. Rev. **81**, 835 (1951); F. S. Goucher, Phys. Rev. **81**, 475 (1951); E. M. Pell, Phys. Rev. **90**, 278 (1953); A. Many, Proc. Phys. Soc. (London) **B67**, 9 (1954).

[2] J. R. Haynes and J. A. Hornbeck, Phys. Rev. **90**, 152 (1953).

[3] W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950).

[4] The xenon spark tube was supplied by H. E. Edgerton of the Massachusetts Institute of Technology, Cambridge, Massachusetts.

pulse breaks down the spark tube and allows the condenser to discharge through it. With $C=500$ mmf the light pulse is about 0.1 μsec long. The pulse was observed using a photomultiplier tube. A lens is used to focus the light on the sample and to allow room between the sample and the spark tube for a copper screen shield. A small amount of shielding is necessary to keep electrical interference from the light source out of the measuring system. A motor driven switch operates the spark coil through a relay. The lowest lifetime measurable with this light source is about 0.2 μsec.

The two end faces of the sample are sandblasted and electrical leads soldered to them. The solder covers the entire end of the sample so that the electric field set up by the measuring current is uniform throughout the sample. A uniform field is required for theoretical analysis of the experiment (see Sec. 3). A glass cryostat is used when the sample temperature is to be varied.

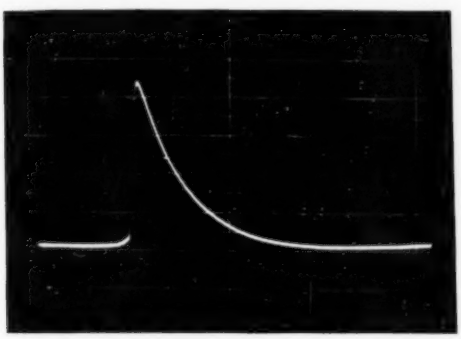


FIG. 3. Typical oscilloscope signal.

An electronic constant current generator is used when currents from 0.1 to 10 milliamperes are required. For larger currents a battery with the appropriate series resistance is needed. The maximum allowable sample current is that for which the electric field begins to cause changes in the observed decay time. This point is discussed in Sec. 3. The photoconductivity pulse is displayed with a Tektronix 512 or 513 oscilloscope, the choice depending on the lifetime being measured. Of course, the frequency response of the detecting system must be adequate to display the decay curve without distortion. In most cases some additional amplification is needed ahead of the oscilloscope.

The lifetime is determined by noting the time required for the amplitude of the decaying signal to drop to $\frac{1}{2}$ its original value. Measurements starting from several different amplitudes should be made to make certain that the decay is truly exponential, i.e., that higher modes are not present (see Sec. 3). Only the center portion of the cathode-ray tube is used. A typical signal observed on the oscilloscope is shown in Fig. 3. The mean lifetime, which is the lifetime usually quoted for semiconductors, is obtained from the half-life as determined from the oscilloscope display by dividing by ln 2. Under normal conditions the sample lifetime can be determined to an accuracy of about 5 percent.

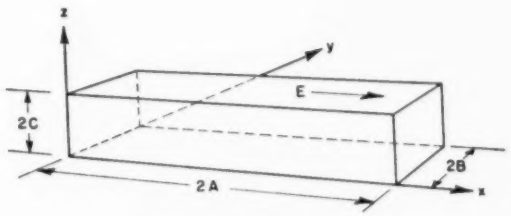


FIG. 4. Sample geometry.

Two other phenomena are sometimes present which may lead to an error in the lifetime determination. The temporary trapping of carriers described by Haynes and Hornbeck[2] results in a long time constant tail on the exponential decay. It is believed that this type of trapping does not occur in germanium at room temperature but it has been observed in germanium at lower temperatures and in silicon at room temperature. For ordinary lifetime measurements the trapping can usually be eliminated by shining a strong continuous light on the sample while the measurement is being made. The second source of possible error is the photovoltaic effect. This is a voltage developed at the end connections of the sample and is present when no current is sent through the sample. The decay time of this voltage is not necessarily the same as the photoconductivity signal. The amplitude of this voltage may be reduced by making good connections to the ends of the sample and by keeping the pulsed light away from the end contacts.

3. THEORY

A. Solutions of the Diffusion Equation

The differential equation governing the excess carrier density n in the interior of a semiconductor with a uniform electric field E_x in the x direction is[3]

$$\frac{\partial n}{\partial t}=-\lambda n-\mu E_x\frac{\partial n}{\partial x}+D\nabla^2 n, \tag{1}$$

where λ = decay probability $=1/\tau_B$ = the reciprocal bulk lifetime, μ = carrier mobility, and D = carrier diffusion constant. The electric field is set up by the current we must send through the sample to measure its conductance. The electric field must be the same throughout the sample if the change in conductance is to be proportional to the total number of added carriers.[3] The samples must therefore be right cylinders with the electric field parallel to the generator of the cylinder. The discussion here will be limited to the case of the rectangular parallelepiped shown in Fig. 4. Some results of the analysis for the circular cylinder are given in the appendix. We assume that the conductivity and the decay probability are uniform throughout the

sample. A solution of (1) is

$$n(x,y,z,t)=G(\cos ax)\exp(\mu E_x x/2D)\cdot(\cos by)(\cos cz)$$
$$\times\exp-\left[\lambda+D(a^2+b^2+c^2)+\frac{(\mu E)^2}{4D}\right]t. \quad (2)$$

If the boundary conditions on the faces of the sample are taken to be

$$D\frac{\partial n_1}{\partial x}=\pm s_a n_1, \quad D\frac{\partial n_1}{\partial y}=\pm s_b n_1, \quad D\frac{\partial n_1}{\partial z}=\pm s_c n_1, \quad (3)$$

where s_a, s_b, and s_c are the so-called surface recombination velocities for the x, y, and z faces, and the values of a, b, and c can be found from the transcendental equations which result when (3) is substituted in (2). Then (2) can be written

$$n_{ijk}(x,y,z,t)=G_{ijk}\left(\cos\frac{\xi_i x}{A}\right)$$
$$\times\exp(\mu E_x x/2D)\cdot\left(\cos\frac{\eta_j y}{B}\right)\left(\cos\frac{\zeta_k z}{C}\right)$$
$$\times\exp-\left[\lambda+D\left(\frac{\xi_i^2}{A^2}+\frac{\eta_j^2}{B^2}+\frac{\zeta_k^2}{C^2}\right)+\frac{(\mu E)^2}{4D}\right]t, \quad (4)$$

where ξ_i, η_j, and ζ_k are, respectively, aA, bB, and cC. Multiple solutions for a, b, and c result from the imposition of the boundary conditions and represent the fundamental and higher modes expected in the solution of this sort of problem. G_{ijk} is the amplitude of the ijk mode. A completely general solution of (1) satisfying any arbitrary spatial distribution of n at $t=0$ can be built up as a sum of solutions of the form (4).

The time dependent term in (4) can be separated into three parts. The first term λ is the decay constant in the bulk and its reciprocal is the so-called bulk lifetime. The second term which we abbreviate by ν_{ijk} is the contribution to the total decay constant arising from diffusion to the walls with subsequent recombination. The third term is the contribution to the total decay constant arising from the additional diffusion current caused by the distortion of the distribution by the electric field. For lifetime measurements this last term should be kept small compared with the sum of the first two, i.e.,

$$\frac{(\mu E)^2}{4D}\ll\lambda+\nu_{ijk}. \quad (5)$$

For example, in an n-type sample where the sample lifetime is 100 μsec the measuring field strength should be much less than 0.25 volt/cm. Keeping this experimental limitation in mind we will omit the field term in our further discussion of (4).

The magnitude of the surface term ν_{ijk} depends on s_a, s_b, and s_c. Shockley[3] has discussed the relation of ν_{ijk} to s for the field free case. We shall discuss three cases here.

Case I: s very large on all faces ($s_a=s_b=s_c=\infty$).

In this case (as also in cases II and III) we assume that the end faces which are sandblasted and covered with solder have a very large surface recombination velocity. If s_b and s_c also are made very large by sandblasting or grinding, then ξ, η, and ζ take the simple form

$$\xi=\frac{i\pi}{2}, \quad \eta=\frac{j\pi}{2}, \quad \zeta=\frac{k\pi}{2},$$

and (4) becomes

$$n_{ijk}(x,y,z,t)=G_{ijk}\sin\frac{i\pi x}{2A}$$
$$\times\exp(\mu E_x x/2D)\cdot\left(\sin\frac{j\pi y}{2B}\right)\left(\sin\frac{k\pi z}{2C}\right)$$
$$\times\exp-\left[\lambda+\frac{\pi^2 D}{4}\left(\frac{i^2}{A^2}+\frac{j^2}{B^2}+\frac{k^2}{C^2}\right)\right]. \quad (6)$$

It is to be noted that this equation is not dependent on the surface recombination velocity. To find the change in conductance of the sample we need the total number of carriers added and hence (6) must be integrated over the volume of the sample. The result is

$$N_{ijk}(t)=G_{ijk}\frac{2A}{i\pi}[(-1)^{i+1}+1]\frac{2B}{j\pi}[(-1)^{j+1}+1]$$
$$\times\frac{2C}{k\pi}[(-1)^{k+1}+1]\exp(-\gamma_{ijk}t), \quad (7)$$

where

$$\gamma_{ijk}=\lambda+\frac{\pi^2 D}{4}\left(\frac{i^2}{A^2}+\frac{j^2}{B^2}+\frac{k^2}{C^2}\right). \quad (8)$$

The reciprocal of γ_{ijk} is the mean life of the excess

TABLE I. Calculated decay constants for the fundamental and some higher modes in an n-type germanium sample 3.0×0.4×0.4 cm with bulk lifetime of 200 μsec, and recombination velocity very high on all surfaces.

Mode	Decay constant $\sec^{-1}$
γ_{111}	10 500
γ_{211}	10 630
γ_{311}	10 860
γ_{411}	11 290
γ_{511}	11 630
$\gamma_{131}=\gamma_{113}$	32 300
γ_{133}	54 000

carrier density for a given mode in the sample. We define the sample lifetime as this lifetime for the lowest mode, $1/\gamma_{111}$. The final time dependence of the excess carrier density $N(t)$ is obtained by integrating over each mode of (6) and adding the results.

$$N(t)=\sum H_{ijk}\exp(-\gamma_{ijk}t). \tag{9}$$

Ordinarily, we want to measure the lifetime corresponding to the first mode only. Amplitudes of the higher modes may be estimated from harmonic analysis of the initial spatial distribution. This information combined with the time constants for the various modes from (8) allows one to estimate the effect of the higher modes. It can be seen from (7) that any mode having one or more even mode numbers makes no contribution to $N(t)$. Higher modes are suppressed by making the initial distribution as nearly like that of the fundamental as possible and, since all higher modes decay faster than the fundamental, by waiting for the higher modes to decay. For example, consider an n-type sample $3.0\times0.4\times0.4$ cm with $1/\lambda=200$ μsec. The decay constants obtained from (8) for some of the lower modes are given in Table I. With this choice of dimensions the modes are of two general types: the $i11$ modes which have nearly the same decay constants as the fundamental, increasing only slowly with increasing i, and the $1jk$ modes all of which have decay constants much larger than that of the fundamental.

The amplitude of the $i11$ modes decreases fairly rapidly as i increases, so that by the time ν_{i11} becomes considerably larger than ν_{111}, the amplitude of the $i11$ modes is so small that they make only a small contribution to the total decay. In this case then, after allowing a short time for the $1jk$ modes to decay, the decay curve observed should be substantially the same as if the fundamental only were present.

Since exact determination of the relative amplitudes of the modes is difficult the decay curve should be studied experimentally to make sure that higher modes are absent. It should be noted here that lifetimes and changes in lifetime obtained by observation of the *amplitude* of the photoconductivity may be in error because of the presence of higher modes.

Case II: s very large on faces perpendicular to x-axis and zero on the other faces ($s_a=\infty$, $s_b=s_c=0$).

The total number of added carriers in this case is

$$N(t)=\sum_{i=1}^{\infty} H_{i00}e^{-\gamma t}, \tag{10}$$

where

$$\gamma=\lambda+\frac{\pi^2Di^2}{4A^2}. \tag{11}$$

All the higher modes for the y and z directions give no contribution to $N(t)$. As previously $N(t)$ is zero if the mode number for the x direction is even. The higher modes may be suppressed in the same way as in the previous case.

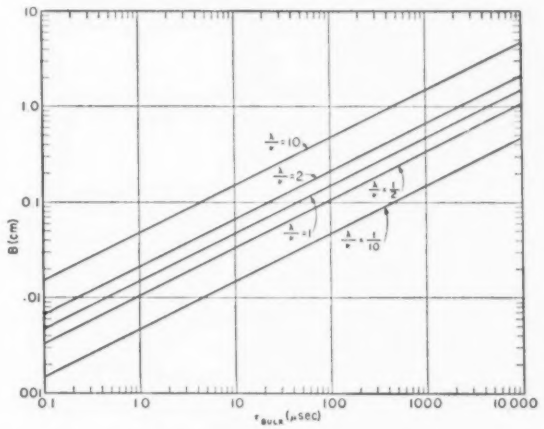


FIG. 5. Sample dimension B vs bulk lifetime for various ratios of the bulk term λ to the surface term ν. This plot is for an n-type sample of square cross section ($2B\times2B$) with all surfaces sandblasted. The diffusion constant was taken as 44 cm²/volt-sec. For bulk lifetime measurements λ/ν should be large. For diffusion constant measurements λ/ν should be small.

Case III: s very large on faces perpendicular to x axis and equal to some intermediate value on the other faces ($s_a=\infty$, $s_b=s_c=s$).

When s on the y and z faces is neither zero nor very large $N(t)$ has properties intermediate between these two limiting cases. For this case if only the fundamental mode is considered $N(t)$ may be written

$$N(t)=He^{-(\lambda+\nu)t}, \tag{12}$$

where ν is the surface contribution to the recombination probability and is a known function[3] of the recombination velocity s, the diffusion constant D, and the sample dimensions. Hence, from measurements of the total recombination probability $\gamma=\lambda+\nu$ under the proper conditions, one can deduce the bulk lifetime, the surface recombination velocity, and the diffusion constant.

B. Applications of the Solutions

(1) *Measurement of Bulk Lifetime*

For measuring the bulk lifetime, ν should obviously be comparable with or smaller than λ. Two measurements can be made: first a measurement with all surfaces sandblasted; the surface term and the resultant bulk lifetime are calculated from (7) and the diffusion constant. A second measurement can be made with a very low-s etch ($s_b=s_c\approx0$) on the surfaces parallel to the measuring field. Assuming that s is very large at the current contacts on the faces perpendicular to the x axis, we may compute the bulk lifetime from (11). In some cases it may be necessary to know s on the side faces (from a measurement on other samples with the same etch) and correct for it. For s very small

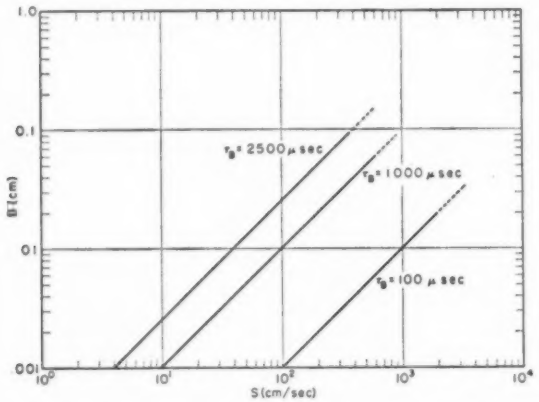


FIG. 6. Sample dimension B *vs* recombination velocity s for various bulk lifetimes for the case where surface term s/B is 10 times λ. The sample is a flat slab of thickness $2B$. Recombination at the edges of the slab is neglected. Dotted lines indicate region where approximation (14) begins to fail.

the loss at the y and z faces gives[3] a contribution to ν of

$$s\left(\frac{1}{B}+\frac{1}{C}\right). \tag{13}$$

From these considerations it is seen that the size of sample necessary for bulk lifetime measurements increases as the bulk lifetime increases (see Fig. 5).

(2) *Measurement of Surface Recombination Velocity*

For measurement of s we must make the surface term predominate by making the sample small in at least one dimension (see Fig. 6). For a sample of small $s(s=s_b=s_e)$ on the y and z faces and with the usual assumption of very large recombination velocity on the x faces, the surface term becomes

$$\nu=\frac{\pi^2 D}{4A^2}+s\left(\frac{1}{B}+\frac{1}{C}\right). \tag{14}$$

It is obvious that the sample should be long compared with one of the transverse dimensions. When s is not so small, the approximation (14) no longer holds and the exact relation between ν and s must be used. Criteria for determining when the various approximations for ν are valid, may be found in reference 3.

(3) *Measurement of the Diffusion Constant and Mobility*

If the bulk term λ is small compared with the surface term ν or if λ is known from a previous measurement, the diffusion constant can be deduced from the measured lifetime from the relation

$$\gamma=\lambda+\frac{\pi^2 D}{4}\left(\frac{1}{A^2}+\frac{1}{B^2}+\frac{1}{C^2}\right). \tag{15}$$

A measurement of the diffusion constant by this method has been reported by Navon, Bray, and Fan.[1] The carrier mobility μ in strongly n or p type material may be deduced from the diffusion constant by use of the Einstein relation

$$\mu=\frac{e}{kT}D, \tag{16}$$

where e is the electronic charge, k is Boltzmann's constant, and T is the absolute temperature. In the intrinsic and near intrinsic range D is related to the diffusion constant for n type material D_n and for p type material D_p by the following relation:[5]

$$D=\frac{n+p}{\dfrac{n}{D_p}+\dfrac{p}{D_n}}, \tag{17}$$

where n and p are the electron and hole densities, respectively. For intrinsic samples $n=p$, and hence $D=2D_pD_n/(D_p+D_n)$. This so-called ambipolar diffusion constant can be measured directly by the present method. The results of such a measurement for temperatures from 300° to 400°K are shown in Fig. 7. These

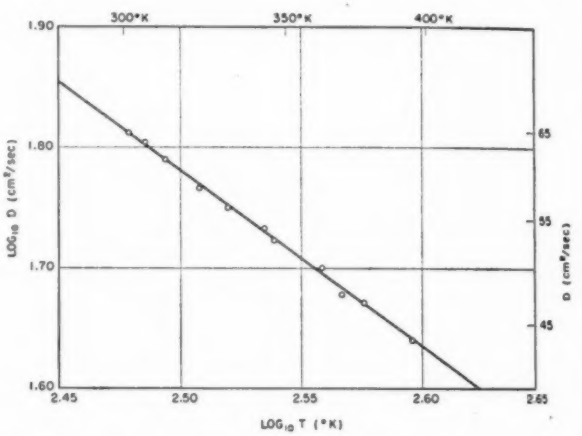


FIG. 7. Ambipolar diffusion constant in intrinsic germanium.

results agree with those calculated using Eqs. (16) and (17) and the drift mobility data of Prince.[6]

The electronic constant current source was designed by Gerald Hitchcox. Dr. S. F. Neustadter obtained the solution of the diffusion equation with the electric field term and the solution for the right circular cylinder. The authors wish to thank many other members of the laboratory, and especially Professor J. E. Thomas, Jr., for their helpful discussions and criticisms.

[5] H. Brooks, Phys. Rev. **90**, 336 (1953).
[6] M. B. Prince, Phys. Rev. **92**, 681 (1953).

APPENDIX. RESULTS FOR A RIGHT CIRCULAR CYLINDER

For a right circular cylinder of radius r and height $2h$ the surface term for small s is

$$\nu=\frac{2s}{r}+\frac{\pi^2 D}{4h^2}. \tag{A1}$$

For large s (sandblasted or ground surface) the surface term is

$$\nu=\frac{\pi^2 D}{4}\left(\frac{9}{4r^2}+\frac{1}{h^2}\right). \tag{A2}$$

In both cases it is assumed that s is very large on the circular end faces. These results are useful if sections of single crystal ingots of circular or near circular cross section are to be measured.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 2 FEBRUARY, 1955

Short-Time Frequency Measurement of Narrow-Band Random Signals by Means of a Zero Counting Process*

HERBERT STEINBERG, PETER M. SCHULTHEISS, CONRAD A. WOGRIN, AND FELIX ZWEIG
Dunham Laboratory of Electrical Engineering, Yale University, New Haven, Connecticut

(Received February 1, 1954)

Any instrumentation for measuring the mean frequency of a narrow-band random signal presents an output quantity which inherently fluctuates about that value which represents the true mean frequency, due to the finite measuring time. In comparing the accuracy of various instrumentations for short-time frequency measurement a useful figure of merit is *variance of the output/(sensitivity)*^{2}. In this paper the figure of merit is obtained for an instrumentation which measures frequency by determining the average number of zero crossings of the signal in a short time. The significant problem is the determination of the mean square number of zeros of a random function in time T. The general result is presented in the form of an integral. For an assumed Gaussian power spectrum the integral is integrated graphically to obtain a figure of merit which is compared with previously published figures of merit of an autocorrelator and frequency discriminator.

INTRODUCTION

A PROBLEM of frequency measurement arises whenever a data transmission system codes the information to be transmitted as a frequency shift of a carrier. Ideally there is a one to one correspondence between frequency and the state of the physical quantity from which the data are obtained. The recovery of the desired information is then readily accomplished by means of any instrumentation for measuring frequency. The problem to be dealt with here is the frequently encountered situation in which the data-taking process is subject to random disturbances. In this case a constant state of the physical quantity results, not in a single frequency, but in a spectrum possibly of the form shown in Fig. 1. If there is no reason to expect a systematic error in either direction, the power spectrum should be symmetrical and the center frequency f_0 can then be identified with the desired data. Any instrumentation for recovering the information in this type of signal has a fundamental limitation; an infinite measuring time is necessary to establish exactly the frequency f_0. In any practical situation where only a finite time is available for the measurement, the accuracy is dependent upon the length of the time interval of the measurement and the width of the power spectrum.

In an earlier paper by the authors[1] two techniques of frequency measurement have been analyzed and compared. They were based respectively on a determination of the short-time autocorrelation by means of an autocorrelator and on the symmetry of the power spectrum by means of what is essentially a frequency discriminator. The present paper discusses a third and somewhat different method of determining frequency, the counting of zeros of the signal in a specified time interval. Since the function being observed is random, the number of zeros counted in any time interval T is itself a

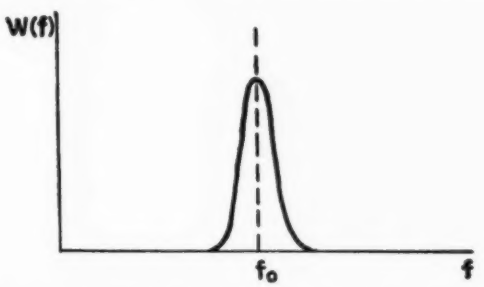


FIG. 1. The power spectrum of a signal whose source is subject to random fluctuations.

* The work described in this paper was performed under contract between General Precision Laboratory, Pleasantville, New York and Dunham Laboratory, Yale University, and was taken from the Eighth Progress Report, SAR/GPL Subcontract No. 34, October 24, 1952.

[1] Schultheiss, Wogrin, and Zweig, J. Appl. Phys. 8, 1025 (1954).

random variable. The useful output then of the frequency measuring device is a voltage proportional to the average number of zeros in T, while the fluctuations about this value represent a disturbance.

In comparing the instrumentations for frequency measurement an intuitively appealing figure of merit is the ratio *variance of the output/(sensitivity)*[2]. A formal justification for its use was given in the paper mentioned above. The sensitivity is defined as the change in dc output for a unit change of center frequency f_0 of the input power spectrum.

The sensitivity of the zero counter is easily obtained by differentiating with respect to f_0 the expression for the expected number of zeros of a random function, an expression which has been derived in the literature.[2] The variance of the output will be derived here in general terms and reduced to the simplest form attainable without assumption of a specific signal correlation function. The final computation of a figure of merit for any specific case then resolves itself into a straight forward numerical integration. This integration has been carried out for a Gaussian power spectrum and indicates a performance roughly comparable to that of a discriminator or autocorrelator adjusted for optimum operation.

The methods used here are only rigorously applicable if the signal dealt with is a stationary process. If, for instance, the frequency f_0 of Fig. 1 were to change with time, the signal is not a stationary process and the power spectrum can no longer be defined, at least in the usual manner. However, if the rate of change of f_0 were very slow as compared to the measuring time interval, and if the fluctuation about f_0 were relatively small, any error introduced by assuming a stationary process would be quite small.

ZEROS OF RANDOM NOISE

Let the noise voltage (x) be represented by the usual Fourier expansion

$$x(t,\sigma)=\sum_{K=1}^{M} a_k(\sigma)\cos\omega_k t+b_k(\sigma)\sin\omega_k t, \qquad (1)$$

where $a_k(\sigma)$ and $b_k(\sigma)$ are random variables of the stochastic process, σ being a variable in abstract measure space.[3] The random voltage has a power spectrum $w(f)$ related to $a_k(\sigma)$ and $b_k(\sigma)$ by the expression

$$\langle\{a_k(\sigma)\}^2\rangle=\langle\{b_k(\sigma)\}^2\rangle=w(f_k)\Delta f. \qquad (2)$$

Let $N(T,\sigma)$ represent the number of zeros of $x(t,\sigma)$ in the interval $[0,T]$. In any practical case the noise voltage has a finite number of maxima and minima and $x(0,\sigma)\neq 0\neq x(T,\sigma)$ with probability 1. From Appendix A it follows that

$$N(T,\sigma)=\lim_{n\to\infty}\int_0^T \delta_n[x(t,\sigma)]\left|\frac{\partial}{\partial t}x(t,\sigma)\right|dt, \qquad (3)$$

where $\lim_{n\to\infty}\delta_n(u)$ is the Dirac delta function.

Equation (3) is intuitively reasonable, for by definition of the δ function

$$\int_{-\epsilon}^{\epsilon}\delta(y)dy=1 \quad \text{for all } \epsilon>0 \qquad (4)$$

so that

$$\int_a^b \delta[f(x)]f'(x)dx=\begin{cases}+1 \text{ if } f(x) \text{ has a simple zero with positive slope in } a,\ b\\ -1 \text{ if } f(x) \text{ has a simple zero with negative slope in } a,\ b\\ 0 \text{ otherwise.}\end{cases}$$

In order to obtain a count of all zeros therefore, the absolute value of the slope is used in Eq. (3).[4,5]

VARIANCE OF THE DISTRIBUTION OF ZEROS

The function $N(T,\sigma)$ is itself a random process and the second moment $\langle N^2(T)\rangle$ of its distribution is given by the expression

$$\langle N^2(T)\rangle=\int_\Omega N^2(T,\sigma)d\sigma \qquad (5)^6$$

where Ω is the entire space of the variable σ. Substituting Eq. (3) into Eq. (5)

$$\langle N^2(T)\rangle=\int_\Omega \lim_{n\to\infty}\left\{\int_0^T \delta_n[x(t,\sigma)]\left|\frac{\partial}{\partial t}x(t,\sigma)\right|dt\right\}^2 d\sigma. \qquad (6)$$

It is easily shown that the integration in σ can be performed before the limiting operation in n since the integrand is uniformly bounded. Hence

$$\langle N^2(T)\rangle=\lim_{n\to\infty}\int_\Omega\left\{\int_0^T \delta_n[x(t,\sigma)]\left|\frac{\partial}{\partial t}x(t,\sigma)\right|dt\right\}^2 d\sigma. \qquad (7)$$

It follows that

$$\langle N^2(T)\rangle=\lim_{n\to\infty}\int_{s=0}^T\int_{t=0}^T\int_\Omega \delta_n[x(t,\sigma)]\delta_n[x(s,\sigma)]\times\left|\frac{\partial}{\partial t}x(t,\sigma)\right|\left|\frac{\partial}{\partial s}x(s,\sigma)\right|d\sigma ds dt. \qquad (8)^7$$

[2] S. O. Rice, Bell System Tech. J. **24**, 51 (1945).

[3] Halmos, *Measure Theory* (D. Van Nostrand Company, Inc., New York, 1950), Chapter IX.

[4] A similar scheme is used by F. L. H. M. Stumpers, "Theory of Frequency-Modulation Noise," Institute of Radio Engineers Sept. 1948, p. 1081. Stumpers employs a device which counts only zeros with positive slope.

[5] A. Kohlenberg, "Notes on the Zero Distribution of Gaussian Noise," Technical Memorandum No. 44, Lincoln Laboratory, Massachusetts Institute of Technology, October 16, 1953. This work uses an almost identical scheme for determining the moments of the zero distribution. It also contains an interesting discussion of possible analytical approaches to the zero distribution problem.

[6] Halmos, see reference 3.

Considerable simplification would result if the limiting operation in n could be carried out before the integrations in t and s. Unfortunately this is not permissible because the integrand would no longer be uniformly bounded. The computation must therefore proceed from Eq. (8), δ_n being represented by some suitable bounded function.

The integration in σ will be performed first. Let

$$\int_{\Omega} \delta_n[x(t,\sigma)]\delta_n[x(s,\sigma)]\left|\frac{\partial}{\partial t}x(t,\sigma)\right| \times\left|\frac{\partial}{\partial s}x(s,\sigma)\right| d\sigma \equiv I_n(s,t). \quad (9)$$

Then

$$I_n(s,t)=\int\int_{-\infty}^{\infty}\int\int \delta_n(x_1)\delta_n(x_4) \times|x_2x_3|\,dF(x_1,x_2,x_3,x_4) \quad (10)^{7}$$

where:

$x_1=x(t,\sigma)$

$x_2=\dfrac{\partial}{\partial t}x(t,\sigma)$

$x_3=\dfrac{\partial}{\partial s}x(s,\sigma)$

$x_4=x(s,\sigma)$

$F(x_1,x_2,x_3,x_4) = P\{x(t)\leq x_1,\ x'(t)\leq x_2,\ x'(s)\leq x_3,\ x(s)\leq x_4\},$

the distribution function for the four variables.

Rice has demonstrated that this distribution function is normal in four dimensions.[8]

$$dF(x_1,x_2,x_3,x_4)=\frac{1}{(2\pi)^2|M|^{\frac{1}{2}}} \times\exp\left[-\frac{1}{2|M|}\sum_{i=1}^{4}\sum_{j=1}^{4}M_{ij}x_ix_j\right] \times dx_1dx_2dx_3dx_4, \quad (11)$$

where M is the moment matrix of the distribution, $|M|$ the corresponding determinant, and M_{ij} the cofactors.

For computational purposes let

$$\delta_n(u)=\frac{n}{(2\pi)^{\frac{1}{2}}}\exp\left[-\frac{n^2u^2}{2}\right]. \quad (12)$$

From Eqs. (10), (11), and (12)

$$I_n(s,t)=\frac{n^2}{(2\pi)^3|M|^{\frac{1}{2}}}\int_{-\infty}^{\infty}dx_2\int_{-\infty}^{\infty}dx_3|x_2x_3| \times\exp\left\{-\frac{1}{2|M|}[M_{22}(x_2^2+x_3^2)+2M_{23}x_2x_3]\right\} \times G(x_2,x_3) \quad (13)$$

where

$$G(x_2,x_3)\equiv\int_{-\infty}^{\infty}dx_1\int_{-\infty}^{\infty}dx_4 \times\exp\left\{-\frac{n^2|M|+M_{11}}{2|M|}(x_1^2+x_4^2) -\frac{1}{|M|}[M_{14}x_1x_4+M_{24}x_2x_4+M_{34}x_3x_4 +M_{13}x_1x_3+M_{12}x_1x_2]\right\} \quad (14)$$

and use has been made of the fact that $M_{22}=M_{33}$, $M_{11}=M_{44}$, and $M_{ij}=M_{ji}$.

Isolating the integration with respect to x_1 in Eq. (14),

$$G(x_2,x_3)=\int_{-\infty}^{\infty}dx_4\exp\left\{-\frac{n^2|M|+M_{11}}{2|M|}x_4^2 -\frac{1}{|M|}[M_{24}x_2x_4+M_{34}x_3x_4]\right\}A(x_2,x_3x_4) \quad (15)$$

where

$$A(x_2,x_3,x_4)\equiv\int_{-\infty}^{\infty}dx_1\exp\left\}-\frac{n^2|M|+M_{11}}{2|M|}x_1^2 -\frac{1}{|M|}[M_{14}x_4+M_{13}x_3+M_{12}x_2]x_1\right\}. \quad (16)$$

From integral tables[9]

$$A(x_2,x_3,x_4)=\frac{|M|^{\frac{1}{2}}(2\pi)^{\frac{1}{2}}}{[n^2|M|+M_{11}]^{\frac{1}{2}}} \times\exp\left\{\frac{[M_{12}x_2+M_{13}x_3+M_{14}x_4]^2}{2|M|[n^2|M|+M_{11}]}\right\}. \quad (17)$$

When Eq. (17) is substituted into Eq. (15), the integral in x_4 becomes identical in form with Eq. (16) and can

[7] Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946), 15.3.2.

[8] S. O. Rice, Bell System Tech. J. (1945).

[9] Bierens de Haan, *Nouvelles Tables D'Integrales Definies* (G. Stechert and Company, 1939), T 28 No. 1.

therefore be evaluated without difficulty. Then

$$G(x_2,x_3)=\frac{2\pi|M|}{\{[n^2|M|+M_{11}]^2-M_{14}^2\}^{\frac{1}{2}}}\exp\left\{\frac{[M_{12}x_2+M_{13}x_3]^2}{2|M|[n^2|M|+M_{11}]}\right.$$
$$\left.+\frac{[(n^2|M|+M_{11})(M_{24}x_2+M_{34}x_3)-M_{14}(M_{12}x_2+M_{13}x_3)]^2}{2|M|[(n^2|M|+M_{11})^2-M_{14}^2][n^2|M|+M_{11}]}\right\}. \quad (18)$$

At this point it becomes convenient to adopt the following definitions:

$$a\equiv\frac{n^2|M|^{\frac{1}{2}}}{4\pi^2[(n^2|M|+M_{11})^2-M_{14}^2]^{\frac{1}{2}}} \quad (19)$$

$$b\equiv\frac{M_{22}}{2|M|}-\frac{M_{12}^2}{2|M|[n^2|M|+M_{11}]}-\frac{[(n^2|M|+M_{11})M_{24}-M_{12}M_{14}]^2}{2|M|(n^2|M|+M_{11})[(n^2|M|+M_{11})^2-M_{14}^2]} \quad (20)$$

$$c\equiv\frac{M_{22}}{2|M|}-\frac{M_{13}^2}{2|M|[n^2|M|+M_{11}]}-\frac{[(n^2|M|+M_{11})M_{34}-M_{13}M_{14}]^2}{2|M|[n^2|M|+M_{11}][(n^2|M|+M_{11})^2-M_{14}^2]} \quad (21)$$

$$d\equiv\frac{M_{23}}{|M|}-\frac{M_{12}M_{13}}{|M|[n^2|M|+M_{11}]}-\frac{[(n^2|M|+M_{11})M_{24}-M_{12}M_{14}][(n^2|M|+M_{11})M_{34}-M_{13}M_{14}]}{|M|[n^2|M|+M_{11}][(n^2|M|+M_{11})^2-M_{14}^2]}. \quad (22)$$

In terms of these parameters the substitution of Eq. (18) into Eq. (13) yields

$$I_n(s,t)=a\int_{-\infty}^{\infty}dx_3\int_{-\infty}^{\infty}dx_2|x_2x_3|$$
$$\times\exp\{-[bx_2^2+cx_3^2+dx_2x_3]\}. \quad (23)$$

The change of variable $u^2\equiv bx_2^2$, $v^2\equiv cx_3^2$ transforms Eq. (23) into

$$I_n(s,t)=\frac{2a}{bc}\int_0^{\infty}\int_0^{\infty}uv\exp[-(u^2+v^2)]$$
$$\times\left\{\exp\left[\frac{d}{(bc)^{\frac{1}{2}}}uv\right]+\exp\left[-\frac{d}{(bc)^{\frac{1}{2}}}uv\right]\right\}dudv. \quad (24)$$

Double integrals of this form have been evaluated by Rice.[10] Using his result

$$I_n(s,t)=\frac{4a}{4bc-d^2}\left[1+\frac{d}{(4bc-d^2)^{\frac{1}{2}}}\tan^{-1}\frac{d}{(4bc-d^2)^{\frac{1}{2}}}\right]. \quad (25)$$

From Eqs. (8) and (9) the desired second moment $\langle N^2(T)\rangle$ is given by

$$\langle N^2(T)\rangle=\lim_{n\to\infty}\int_0^T\int_0^T I_n(s,t)dsdt. \quad (26)$$

The integration may be simplified by recognizing that all nonconstant terms in the moment matrix M are correlation functions (or their derivatives) for the interval $\tau=t-s$. Hence $I_n(s,t)$ is a function of the single variable $\tau=t-s$.

$$\langle N^2(T)\rangle=\lim_{n\to\infty}\int_0^T\int_0^T I_n(t-s)dsdt. \quad (27)$$

The double integral can be reduced to a single integral by the change of variable $\tau\equiv t-s$ and $\rho\equiv t+s$. The integration in ρ is readily performed leaving

$$\langle N^2(T)\rangle=\lim_{n\to\infty}\int_{-T}^{T}d\tau(T-|\tau|)I_n(\tau). \quad (28)$$

Finally it can be seen from an examination of the moment matrix [Eq. (35)] and Eqs. (19)–(22) that $I_n(\tau)$ is an even function of τ so that Eq. (33) can be reduced to the expression

$$N^2(T)=\lim_{n\to\infty}2\int_0^T d\tau(T-\tau)I_n(\tau). \quad (29)$$

$I_n(\tau)$ is given by Eq. (26) and the required cofactors M_{ij} may be obtained from the moment matrix

$$M=\begin{vmatrix}\psi_0 & 0 & \psi_\tau' & \psi_\tau\\ 0 & -\psi_0'' & -\psi_\tau'' & -\psi_\tau'\\ \psi_\tau' & -\psi_\tau'' & -\psi_0'' & 0\\ \psi_\tau & -\psi_\tau' & 0 & \psi_0\end{vmatrix}. \quad (30)$$

Computation from Eq. (29) would be quite prohibitive because of the complexity of Eqs. (19)–(23). Considerable simplification would result if the limiting operation in n could be carried out prior to integration. As pointed out earlier, this is not possible, but it will be observed that $I_n(\tau)$ does, indeed, converge uniformly everywhere

[10] S. O. Rice, see reference 2, p. 69, Eq. (3.5–4), and derivation pp. 66–67.

except at the origin. The integration may therefore be broken up into two parts.

$$\langle N^2(T)\rangle=\lim_{n\to\infty} 2\int_0^{\epsilon} d\tau(T-\tau)I_n(\tau) +2\int_{\epsilon}^{T}\lim_{n\to\infty}(T-\tau)I_n(\tau)d\tau, \quad (31)$$

where ϵ is an arbitrarily small number. Now from Eqs. (25) and (19)–(23)

$$I(\tau)\equiv\lim_{n\to\infty} I_n(\tau)=\frac{|M|^{\frac{1}{2}}}{\pi^2(M_{22}^2-M_{23}^2)} \times\left[1+\frac{M_{23}}{(M_{22}^2-M_{23}^2)^{\frac{1}{2}}}\tan^{-1}\frac{M_{23}}{(M_{22}^2-M_{23}^2)^{\frac{1}{2}}}\right]. \quad (32)$$

If one recognizes that

$$|M|=\frac{M_{22}^2-M_{23}^2}{\psi_0^2-\psi_\tau^2} \quad (33)$$

Eq. (32) assumes the form

$$I(\tau)=\frac{(M_{22}^2-M_{23}^2)^{\frac{1}{2}}}{\pi^2(\psi_0^2-\psi_\tau^2)^{\frac{1}{2}}} \times\left[1+\frac{M_{23}}{(M_{22}^2-M_{23}^2)^{\frac{1}{2}}}\tan^{-1}\frac{M_{23}}{(M_{22}^2-M_{23}^2)^{\frac{1}{2}}}\right]. \quad (34)$$

Equation (34) differs from the expression derived by Rice[11] for the probability density of a zero at $t=\tau$ given a zero at the origin only by the constant multiplier $1/\pi[-\psi_0''/\psi_0]^{\frac{1}{2}}$ which is, of course, simply the expected number of zeros per second. The second integration in Eq. (31) now no longer presents any obstacles to graphical evaluation although the computational work remains fairly formidable.

The first term in Eq. (31) may be handled in the following manner. The number of zeros in the interval $0\leq t\leq\epsilon$ is represented by $N(\epsilon,\sigma)$ in accordance with the usage adopted in Eq. (5). The expected number of zeros in the interval is known from Rice's work.[12]

$$\langle N(\epsilon)\rangle=\int_\Omega N(\epsilon,\sigma)d\sigma=\frac{1}{\pi}\left[-\frac{\psi_0''}{\psi_0}\right]^{\frac{1}{2}}\epsilon. \quad (35)$$

As $\epsilon\to 0$ the probability that more than one zero occurs in the interval $0\leq t\leq\epsilon$ becomes vanishingly small and one need only consider the cases $N(\epsilon,\sigma)=0$ and $N(\epsilon,\sigma)=1$. Under these circumstances the mean square value is clearly equal to the mean

$$\langle N^2(\epsilon)\rangle=\int_\Omega N^2(\epsilon,\sigma)d\sigma=\frac{1}{\pi}\left[-\frac{\psi_0''}{\psi_0}\right]^{\frac{1}{2}}\epsilon \quad (36)$$

since $N(\epsilon,\sigma)=N^2(\epsilon,\sigma)$ for all permissible values of that random variable.

Identifying Eq. (36) with Eq. (29) for $T=\epsilon$ one obtains

$$\frac{1}{\pi}\left[-\frac{\psi_0''}{\psi_0}\right]^{\frac{1}{2}}\epsilon=2\lim_{n\to\infty}\int_0^{\epsilon}(\epsilon-\tau)I_n(\tau)d\tau. \quad (37)$$

Differentiating with respect to the parameter ϵ

$$\frac{1}{\pi}\left[-\frac{\psi_0''}{\psi_1}\right]^{\frac{1}{2}}=2\lim_{n\to\infty}\int_0^{\epsilon} I_n(\tau)d\tau. \quad (38)$$

Equation (38) could be interpreted by noting that $I_n(\tau)$ must approach a Dirac delta function at 0^+ as $n\to\infty$

$$\lim_{n\to\infty} I_n(\tau)=\frac{1}{2\pi}\left[-\frac{\psi_0''}{\psi_0}\right]^{\frac{1}{2}}\delta^+(\tau) \text{ for } \tau \text{ near zero.} \quad (39)$$

With this interpretation the limiting operations in Eq. (29) could be interchanged and the integration is then essentially that of Rice's function[11] over the entire interval $0\leq t\leq T$. The δ function is of no interest in Rice's work but it is, in fact, implied by the assumption of a zero at the origin. By this reasoning, or by direct use of Eq. (43) in Eq. (36), the mean square number of zeros in time T can be written in the form

$$\langle N^2(T)\rangle=\frac{1}{\pi}\left[-\frac{\psi_0''}{\psi_0}\right]^{\frac{1}{2}}T+\frac{2}{\pi^2}\int_0^T d\tau(T-\tau) \times\frac{(M_{22}^2-M_{23}^2)^{\frac{1}{2}}}{(\psi_0^2-\psi_\tau^2)^{\frac{1}{2}}}\left[1+\frac{M_{23}}{(M_{22}^2-M_{23}^2)^{\frac{1}{2}}} \times\tan^{-1}\frac{M_{23}}{(M_{22}^2-M_{23}^2)^{\frac{1}{2}}}\right]. \quad (40)$$

Equation (40) gives the desired result in the simplest form attainable without assumption of a specific signal power spectrum or correlation function. From this point on computation must proceed on a numerical basis, for the substitution of any reasonable algebraic expression for ψ_τ leads to an integrand of sufficient complexity to make formal evaluation almost certainly hopeless.

COMPUTATION FOR AN ASSUMED POWER SPECTRUM

In order to facilitate comparison with earlier work on the autocorrelator and the discriminator, the following signal power spectrum was used in numerical computations

$$w(f)=\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\exp\left[-\frac{(f-f_0)^2}{2\sigma^2}\right] \quad (41)$$

where $\sigma=0.15f_0$. The function $I(\tau)$ corresponding to

[11] S. O. Rice, see reference 3, p. 60, Eq. (3.4–10).
[12] S. O. Rice, see reference 3, p. 54, Eq. (3.3–10).

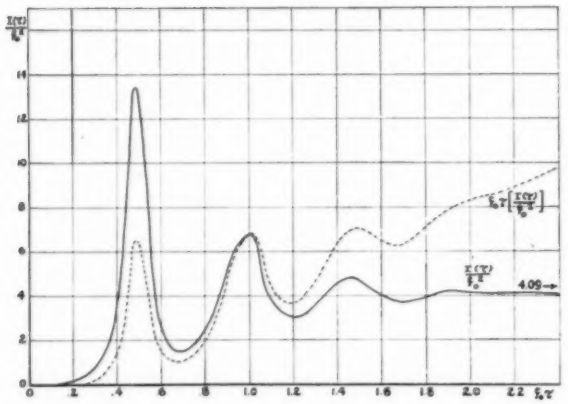


FIG. 2. The fluctuations $I(\tau)$ and $\tau I(\tau)$ of Eq. (34) for the Gaussian power spectrum with standard deviation $\sigma=0.15f_0$ and normalized with respect to f_0. The integral of Eq. (42) is evaluated graphically from these curves.

this power spectrum is shown plotted in Fig. 2. Note that the abscissa and ordinate have been normalized with respect to f_0 and $1/f_0^2$, respectively. The integration

$$\langle N^2(T)\rangle=\langle N(T)\rangle+2\int_0^T(T-\tau)I(\tau)d\tau$$

$$=\langle N(T)\rangle+2\int_0^{f_0T}(f_0T-f_0\tau)\frac{I(\tau)}{f_0^2}d(f_0\tau) \quad (42)$$

is now easily carried out by graphical means. It is convenient to use the fact that $I(\tau)$ has essentially reached its final value $1/\pi^2[-\psi_0''/\psi_0]$, the square of the expected number of zeros/sec, when $f_0\tau=2.5$. Thus the graphical procedure need only cover the range $0\leq f_0\tau\leq 2.5$. The numerical result, the mean square number of zeros in time T, is given by the expression

$$\langle N^2(T)\rangle=4.09f_0^2T^2+0.268f_0T$$
$$=\langle N(T)\rangle^2+0.268f_0T. \quad (43)$$

Substracting the square of the expected number of zeros, $\langle N(T)\rangle^2=4.09f_0^2T^2$, one obtains the desired variance

$$\mu_{11}\equiv\langle N^2(T)\rangle-\langle N(T)\rangle^2=0.268f_0T. \quad (44)$$

The dc sensitivity of a device counting zeros in the manner discussed is given by the relation

$$S\equiv\frac{d}{df_0}\langle(N(T))\rangle=2.024T. \quad (45)$$

Hence the figure of merit, the ratio of the variance to the square of the sensitivity, reduces to the following form

$$\frac{\mu_{11}}{S^2f_0^2}=\frac{0.268f_0T}{4.09f_0^2T^2}=0.00985\frac{1/T}{\sigma}. \quad (46)$$

For the purpose of general comparison, Eq. (46) has been normalized with respect to f_0^2. The corresponding figures for the discriminator and autocorrelator instrumentations derived in the previous paper are

$$\text{Discriminator}\ \frac{\mu_{11}}{S^2f_0^2}=0.0079\frac{\sigma_0}{\sigma} \quad (47)$$

$$\text{Autocorrelator}\ \frac{\mu_{11}}{S^2f_0^2}=0.00825\frac{\sigma_0}{\sigma}. \quad (48)$$

In these two systems the smoothing operation was assumed to be accomplished by a low pass filter with Gaussian characteristic and standard deviation σ_0. The zero counting analysis, on the other hand, postulates a device which perfectly retains information for T seconds and then discards it entirely and instantaneously. It would therefore be incorrect to identify $1/T$ exactly with the band width σ_0 of an equivalent low pass filter.

One can obtain a somewhat better basis of comparison by considering the Gaussian filters of the descriminator and autocorrelator replaced by low pass filters with the weighting function

$$w(t)=u(t)-u(t-T) \quad (49)$$

where

$$u(t)\equiv\begin{cases}0 & \text{for } t<0\\ 1 & \text{for } t\geq 0.\end{cases} \quad (50)$$

This is equivalent to a transfer function $w(\omega)$ given by the expression

$$w(\omega)=\frac{1-e^{-j\omega T}}{j\omega} \quad (51)$$

where $w(\omega)$ is the frequency response function of the filter. When used in conjunction with an autocorrelator, such a filter produces the approximate figure of merit

$$\frac{\mu_{11}}{S^2f_0^2}=0.0033\frac{1/T}{\sigma}. \quad (52)$$

Equation (52) differs from Eq. (46) by a factor 3 with the advantage on the side of the autocorrelator. It should, however, be recalled that the figure of merit measures the square of the rms fluctuation so that the discrepancy is certainly not one in orders of magnitude and one would probably be justified in calling the performance comparable.

CONCLUSION

An analysis has been carried out to compare the performance of a frequency sensitive device utilizing a zero counting method with alternative instrumentations such as a discriminator or autocorrelator. The mean square number of zeros of a random voltage has been evaluated in general terms and specific numerical computations have been made for a Gaussian signal

spectrum. On the basis of these calculations, the zero counting instrumentation appears to be roughly comparable in performance with a discriminator or autocorrelator. The corresponding computation for a signal in the presence of wide band noise offers no theoretical difficulties but has not been carried out because of the extensive numerical work involved.

ACKNOWLEDGMENT

The work described in this paper was done under contract to General Precision Laboratory, Pleasantville, New York, and was indirectly sponsored by the United States Air Force. The authors wish to acknowledge particularly the help of Dr. F. B. Berger of the General Precision Laboratory who contributed many valuable suggestions in the course of this work.

APPENDIX A

Theorem: Let $f(t)$ be an absolutely continuous function on $[0,T]$ such that $f(0)\neq 0$ and $f(T)\neq 0$. Then $f(t)$ shall have N zeros $(0<N<\infty)$ in the interval. Thus $f'(t)$ has the following properties:

1. $f'(t+0)$ and $f'(t-0)$ exist at each t in $[0,T]$ (at the end point read only the inside limits).
2. $f'(t+0)\cdot f'(t-0)\leq 0$ for only a finite number of t in $[0,T]$.

Let $\{\delta_n(\mu)\}$ be defined such that

(a) $\delta_n(\mu)=\delta_n(-\mu)\geq 0 \quad -\infty<\mu<\infty$

(b) $\int_{-\infty}^{\infty}\delta_n(\mu)d\mu=1$

(c) an $\epsilon>0$ implies that

$$\lim_{n\to\infty}\int_{-\epsilon}^{\epsilon}\delta_n(\mu)d\mu=1.$$

Then, the number of zeros of the function $f(t)$ in $[0,T]$ is given by

$$N=\lim_{n\to\infty}\int_0^T\delta_n[f(t)]\,|f'(t)|\,dt. \tag{1A}$$

Proof: Let S be the set of points

$$S=\{t\,|\,t=0,\quad t=T,\quad f'(t+0)f'(t-0)\leq 0\}-\{t\,|\,f(t)=0\}.$$

Thus S is a finite set by assumption (2) above.

Let: $M=$ number of elements of S and let

$$2\epsilon_0=\min_{t\in S}|f(t)|.$$

For $\gamma>0$ there exists an $n(\gamma)$ such that for an $n\geq n(\gamma)$

$$\int_{-\epsilon_0}^{\epsilon_0}\delta_n(\mu)d\mu>1-\frac{\gamma}{2M}.$$

Let: $\{t_k\}_{k=1}^{N}$ be the set of zeros of $f(t)$ and let

$$a_k=\sup\{t\,|\,t<t_k,\,|f(t)|=\epsilon_0\}$$
$$b_k=\inf\{t\,|\,t>t_k,\,|f(t)|=\epsilon_0\}.$$

Then in the intervals $a_k\leq t\leq t_k$ and $t_k\leq t\leq b_k$, $f(t)$ is strictly increasing or decreasing and $f'(t)\geq 0$ or $f'(t)\leq 0$. Thus

$$\int_{t_k}^{b_k}\delta_n[f(t)]\,|f'(t)|\,dt = \int_{a_k}^{t_k}\delta_n[f(t)]\,|f'(t)|\,dt=\int_0^{\epsilon_0}\delta_n(\mu)d\mu$$

Now for $f'(t)\geq 0$, $f(b_k)=\epsilon_0$ and $f(a_k)=-\epsilon_0$ and

$$\int_{t_k}^{b_k}\delta_n[f(t)]f'(t)dt=\int_0^{\epsilon_0}\delta_n(\mu)d\mu$$

$$\int_{a_k}^{t_k}\delta_n[f(t)]f'(t)dt=\int_{-\epsilon_0}^{0}\delta_n(\mu)d\mu$$

and for $f'(t)\leq 0$, $f(b_k)=-\epsilon_0$ and $f(a_k)=\epsilon_0$ and

$$-\int_{t_k}^{b_k}\delta_n[f(t)]f'(t)dt=-\int_0^{-\epsilon_0}\delta_n(\mu)d\mu$$

$$-\int_{a_k}^{t_k}\delta_n[f(t)]f'(t)dt=-\int_{\epsilon_0}^{0}\delta_n(\mu)d\mu.$$

Therefore

$$\sum_1^N\int_{a_k}^{b_k}\delta_n[f(t)]\,|f'(t)|\,dt=2N\int_0^{\epsilon_0}\delta_n(\mu)d\mu.$$

By using the points of the set S as division points, computation shows

$$0\leq\int_0^{a_1}\delta_n[f(t)]\,|f'(t)|\,dt+\int_{b_N}^{T}\delta_n[f(t)]\,|f'(t)|\,dt + \sum_1^{N-1}\int_{b_k}^{a_{k+1}}\delta_n[f(t)]\,|f'(t)|\,\delta(t) \leq(N+M)\int_{|\mu|\geq\epsilon_0}\delta_n(\mu)d\mu.$$

Hence $n\geq n(\gamma)$ implies

$$\lim_{n\to\infty}\int_0^T\delta_n[f(t)]\,|f'(t)|\,dt\leq N+(N+M)\frac{\gamma}{2M} \leq N+\gamma\geq 2N\left(\frac{1}{2}-\frac{\gamma}{4M}\right)\geq N-\gamma$$

where γ is arbitrarily small. Thus the theorem of Eq. (1A) is proved.

Internal Friction of Metal Single Crystals

JOHANNES WEERTMAN
Naval Research Laboratory, Washington 25, D. C.
(Received April 5, 1954)

Semiquantitative calculations were made of that portion of the internal friction of annealed and moderately cold-worked metal single crystals which is due to dislocation motion. The calculations were made on the basis of a qualitative theory of Nowick. It is shown that Nowick's theory leads to reasonable orders of magnitude of the observed internal friction and changes of the modulus of elasticity. At low temperature or at high frequencies (megacycle range) Nowick's hysteresis mechanism becomes inoperative and theories of Koehler and of Eshelby have to be considered to calculate the internal friction.

INTRODUCTION

IN a recent article Nowick[1] has explained qualitatively the internal friction of cold-worked metal single crystals. We believe that his over-all picture is correct. It is the purpose of this article to augment part of Nowick's qualitative picture with some semiquantitative calculations based on dislocation models to show that his picture leads to reasonable numerical values of the decrement and changes in the dynamic modulus. In the first part of the paper we will limit the discussion to the case where the frequency of the oscillatory stress lies in the kilocycle or lower frequency range. The last section will concern the megacycle range of frequency where a theory of Koehler[2] should be valid.

Nowick has come to the conclusion that the internal friction of cold-worked metals is a phenomenon which can be divided into three effects. The first effect, characteristic of annealed and slightly cold-worked metals, is frequency independent but amplitude dependent and can be eliminated by moderate cold working. Nowick calls this the "nonlinear" effect. Nowick explains this effect as being due to a hysteresis phenomenon in which dislocations are moved from one low energy position to another. Weertman and Salkovitz,[3] however, have presented a dislocation hysteresis model which can give rise to an internal friction which is independent of strain amplitude of vibration. The term "nonlinear" is therefore too restrictive and the term "hysteresis" is perhaps better to characterize the first effect.

The second effect Nowick calls a "Köster" effect and it is used to describe the high internal friction which is observed after moderate cold working of a crystal. This internal friction is rather unstable and decays rapidly under low temperature annealing. The origin of this internal friction he pictures as due to the motion of the loosely bound dislocations in the interior of subgrains produced by cold working. These dislocations can move to the tighter binding positions at the subgrain boundaries rather easily and there no longer contribute so much to the energy loss.

The third effect, which Nowick calls the "viscosity" effect, occurs in severely cold-worked materials and is due to the motion of dislocations in the subgrain boundries. This motion is enhanced by raising the temperature, making the internal friction strongly dependent on the temperature and the frequency. This third effect we will not discuss.

ANNEALED AND SLIGHTLY COLD-WORKED METALS IN THE KILOCYCLE OR LOWER FREQUENCY RANGE

The mechanism which accounts for the internal friction of metals in this frequency range and state of the metal was suggested by Nowick[1,4,5] as due to hysteretic motion of dislocation lines. In a paper by Weertman and Salkovitz[3] a specific hysteresis model was set up which can be used to estimate the magnitude of the internal friction and changes in the Young's modulus. This model was applied to the case of the internal friction of lead but it can be used to account for the internal friction of other metals. Briefly, the model is this: In a lattice containing randomly spaced impurity atoms which do not have too high a binding energy to dislocation lines, the dislocation feels a stress field due to the impurity atoms of average wavelength $\lambda=b/c^{\frac{1}{2}}$, where b is the length of the Burger's vector and c is the fraction of the total number of atoms which are impurity atoms. The amplitude of the stress field varies from zero to very large values. The average value of the stress amplitude which is given by the theories of Mott and Nabarro[6–8] is

$$\sigma_M=(2\pi)^{-1}\mu\epsilon^{4/3}c^{11/9}[\ln 1/c]^{4/3}, \quad (1)$$

where μ is the shear modulus and ϵ the fractional difference in the size of the solute and solvent atoms. As the dislocation moves through the crystal the number of potential hills traversed having values in the region σ to $\sigma+d\sigma$ will be proportional to $e^{-\sigma^2/\sigma_M^2}d\sigma$. This means that there are very few places in the crystal where the

[1] A. S. Nowick, J. Appl. Phys. **25**, 1129 (1954).
[2] J. S. Koehler, *Imperfections in Nearly Perfect Crystals* (John Wiley & Sons, Inc., New York, 1952), p. 197.
[3] J. Weertman and E. I. Salkovitz, Acta Metallurgica (to be published).
[4] A. S. Nowick, Metal Progr. **4**, 1 (1953).
[5] A. S. Nowick, Phys. Rev. **80**, 249 (1950).
[6] F. R. N. Nabarro, Proc. Phys. Soc. (London) **58**, 669 (1946).
[7] N. F. Mott and F. R. N. Nabarro, *Report of a Conference on the Strength of Solids* (Physical Society, London, 1948), p. 1.
[8] N. F. Mott, *Imperfections in Nearly Perfect Crystals* (John Wiley & Sons, Inc., New York, 1952), p. 173.

amplitude of the stress field is much greater than σ_M. Under an alternating stress $\sigma\cos(2\pi\nu t)$ dislocations in potential valleys surrounded by hills of height less than σ will move back and forth over these hills in an hysteretic fashion. The energy lost by a dislocation line segment per cycle will be equal to the distance it moves, approximately equal to λ, times the force acting on it at the moment when the internal stresses are no longer capable of preventing it from moving over the potential hill (see appendix I). The total energy lost per cycle will be the sum of the energy losses of all the line segments which are able to move over potential hills under the applied stress. The total energy loss is thus equal to the integral of the product of the length of dislocation lines lying next to potential hills of stress amplitude height σ to $\sigma+d\sigma$ times $\sigma b\lambda$, integrated from 0 to σ. This integral is approximately equal to

$$\frac{N}{\sigma_M}\lambda b\int_0^{\sigma}\sigma d\sigma=\frac{N\lambda\sigma^2 b}{2\sigma_M}$$

where N is the total length of the dislocation lines per unit volume.

If σ is the longitudinal stress of a specimen undergoing longitudinal vibration, account has to be taken of the fact that the stress on the slip planes will be less than σ. For a f.c.c. metal the dislocations will be distributed on four slip planes having three different Burger's vectors on each plane. Taking this distribution into account, the total energy loss is approximately equal to $0.02N\lambda\sigma^2 b/\sigma_M$. The decrement will thus be equal to

$$\text{decrement}=\frac{\text{energy lost/cycle}}{\text{twice stored energy}}\sim\frac{0.02N\lambda b\mu}{\sigma_M} \qquad (2)$$

when the applied stress is less than σ_M. The decrement is independent of frequency and of stress amplitude up to a stress approximately equal to σ_M. It should increase with stress amplitude at greater stresses.

Along with internal friction the dynamic modulus should also be discussed. What is of interest, actually, are changes of the modulus, since only changes will indicate what part of the total strain is definitely due to something like dislocation motion.

The additional strain produced by dislocations moving from one potential valley to another for stresses smaller than σ_M is of the order of $0.02Nb\lambda\sigma/\sigma_M$. If μ is the true modulus of a dislocation free material and $\bar{\mu}$ is the effective dynamic modulus, then

$$\frac{\mu-\bar{\mu}}{\mu}\sim\frac{0.02b\lambda N\mu}{\sigma_M}. \qquad (3)$$

Equations (3) and (2) are thus equal. It is an experimental fact[2] that fractional changes in the modulus (due to, say, annealing or slight cold working) are about of the same order of magnitude as changes in the decrement.

TABLE I.

Material	Approximate room temp. decrement at low strain amplitude and in annealed condition	Purity
Cu[a,j]	3×10^{-5}	99.998% to 99.99%
Cu[b]	3×10^{-5}	99.985%
Cu[l,c]	5×10^{-4}	99.999%
Al[b]	10^{-4}	99.975%
Al[d]	$10^{-5}-10^{-4}$	99.99%
Brass[e]	10^{-5}	
Mg[d]	10^{-5}	99.96%
Pb[k]	10^{-2}	99.997%
Pb[k]	2×10^{-3}	99.9 to 99.99%
Zn[f,g]	10^{-4}	99.99%
Zn[h]	10^{-3}	99.996%
Zn[i]	10^{-5} to 10^{-4}	99.999%
Zn[a]	10^{-5}	99.999%

[a] T. A. Read, Trans. Am. Inst. Mining Met. Engrs. **143**, 30 (1941).
[b] G. H. Found, Trans. Am. Inst. Mining Met. Engrs. **161**, 120 (1945).
[c] J. Weertmann and J. S. Koehler, J. Appl. Phys. **24**, 624 (1953).
[d] J. Pittenger, thesis, Carnegie Institute of Technology (1951).
[e] Zener, Clarke, and Smith, Trans. Am. Inst. Mining Met. Engrs. **147**, 90 (1942).
[f] I. H. Swift and J. E. Richardson, J. Appl. Phys. **18**, 417 (1947).
[g] T. A. Read and E. P. T. Tyndall, J. Appl. Phys. **17**, 713 (1946).
[h] S. Takahashi, J. Appl. Phys. **23**, 866 (1952).
[i] E. I. Salkovitz, unpublished data.
[j] See reference 5.
[k] See reference 3.
[l] See reference 10.

Since Eq. (1) was calculated by Mott on the basis of the length of the smallest segment of dislocation line which can be moved as a unit without causing movement in the remaining length of the dislocation, this equation and, thus, Eqs. (2) and (3) break down when this length is greater than the length of a Frank-Read source.

Equations (2) and (3) should hold for the case of lead in the impurity range of $c=10^{-4}$ to 10^{-2}. For a value of $N=10^8/\text{cm}^2$ [9] and $\epsilon=0.1$ the above values of c substituted into Eq. (2) give the values 7×10^{-3} and 1.3×10^{-5} for the decrement. In the case of lead at room temperature for this impurity range[3] a typical value of the decrement is 2×10^{-3}.

The predicted decrease of the internal friction with increasing impurity content is observed in lead for temperatures above 100°C. At room temperature the situation for lead is not clear.[3] For copper, workers using 99.999 percent pure stock find a decrement which is an order of magnitude higher than by those using 99.999 percent or less pure material (see Table I). Marx and Koehler[10] have also shown that annealing copper crystals in a hydrogen atmosphere results in a smaller decrement than a corresponding anneal in a vacuum. Takahashi* and Konno have observed a decrease of internal friction with increasing alloying content in dilute alloys of copper. The decrease is not as rapid as that predicted by Eq. (2). They have also observed that

[9] A. H. Cottrell, *Dislocations and Plastic Flow in Crystals* (Clarendon Press, Oxford, 1953).
[10] J. Marx and J. S. Koehler, Pittsburgh Conference on the Plastic Deformation of Cryst. Solids (1950), p. 171.
* Private communication.

the break in the internal friction-strain amplitude curves occurs at higher stresses for higher alloying content in agreement with the results of Weertman and Salkovitz on lead alloys. Their values of the stress at the break when plotted as a function of σ_M (given by Eq. (1)) falls on the same curve as do the lead alloy values.

It should be emphasized that Eqs. (2) and (3) are applicable only if the impurities are distributed at random throughout the crystal lattice. In addition to this requirement the dislocation should not be pinned down by the impurity atoms. Simple calculations will show when these conditions can be met.

If the impurity atoms have time enough to reach their equilibrium positions, the concentration of impurities near a dislocation line will be given by $c=c_0\exp(Q/kT)$ where c is the actual concentration, c_0 the average, and Q the binding energy.[9] Table II gives representative values of c/c_0 at 300°K for some metals assuming that the impurity has an $\epsilon=0.1$. From these values it can be seen that lead comes closest to meeting the requirement that there be no segregation of impurities at the dislocation lines. For copper and zinc, for example, there will be segregation under annealing conditions where the impurity atoms can diffuse to the dislocation lines. The result of this segregation would be to increase the value of σ_M and decrease the internal friction.

In high purity zinc (99.999 percent) very low values of the decrement (10^{-5}) are reported. A good example of the fact that σ_M is increased by annealing in the case of zinc is given in the paper by Swift and Richardson (see Table I, reference f). They show that the strain amplitude at which a rapid rise in the internal friction occurs increases with increasing annealing time at room temperature. The decrement decreased with time as long as the annealing was done while the specimen remained mounted in the experimental unit.

There is one method to get around the segregation of impurities at the dislocation lines. If a dislocation line is suddenly moved into a new region of the crystal and the temperature is such that diffusion will occur only very slowly, that dislocation, under an oscillatory stress, will lose energy as calculated by Eq. (2). It is well known that crystals of zinc and copper are extremely sensitive to handling. Their decrement will increase a factor of 10 or 100 under applied loads well below the yield point or even under careful handling when no loads are applied. It does not seem possible that the yield point in the crystal is exceeded in such handling so as to increase the number of dislocations. It is proposed here that the cause of this sensitivity is that under these small stresses, which are lower than the yield point, a great many dislocations are moved from their tight binding positions created by the segregated impurities into virgin territory where Eq. (2) can be used to calculate their energy loss under oscillatory stresses. For small loads the decrement of copper[5,10–14] is observed to increase by a factor of 10 to 100 which gives values of the decrement ranging from 10^{-4} to 10^{-3}. These values are of the order predicted by Eq. (2) for a value of N equal to $10^8/\text{cm}^2$.

TABLE II. Equilibrium values of c/c_0 at 300°K assuming that $\epsilon=0.1$.

Metal	c/c_0	μ[a]	b
Pb	13	0.50×10^{11} dynes/cm^2	3.5×10^{-8} cm
Cu	360	3.2	2.55
Ag	170	2.0	2.88
Al	590	2.5	2.86
Zn	4700	4.0	2.66

[a] Shear modulus across slip plane calculated from data given by R. F. S Hearmon, Revs. Modern Phys. **18**, 409 (1946).

The second condition for Eqs. (2) and (3) to hold is that the dislocations should not be pinned down by impurity atoms. For this condition to hold it is only necessary that the period of the oscillatory stress be larger than the time required for a dislocation to break away from a pinning atom by thermal fluctuations. If ν is the frequency of vibration, the following inequality must hold:

$$\nu<10^{13}e^{-Q/kT}$$

where 10^{13} is the frequency of vibration of a dislocation line. This inequality becomes an equality at the temperature where

$$\ln\nu=13\ln10-Q/kT.$$

Table III gives values of the temperature where this last relationship holds true for an $\epsilon=0.1$. Equations (2) and (3) will be valid for temperatures greater than these if there is no segregation.

At temperatures lower than those given in Table III the dislocations will be pinned down here and there by impurity atoms. Mott[8] has shown that if the impurities are distributed at random the dislocations will assume a zigzag form with a distance of $2b/[c^{2/3}\epsilon^{1/3}]$ between binding points. The stress required to break the dislocations away from the pinning impurities Mott calculates to be of the order of $\sigma_M^*=\frac{1}{8}\mu\epsilon^{4/3}c^{2/3}$.

Koehler[2] has shown that the length of dislocation line having loop lengths greater than the average distance between pinning points (in this case equal to $2b/[c^{2/3}\epsilon^{1/3}]$) decreases exponentially with increasing loop length. Therefore up to the stress σ_M^* there should be very little hysteresis taking place. Koehler's theory should be applicable then. In the last section and in the appendix the magnitude of the decrement is worked out for this case. In the kilocycle frequency range it should be negligibly small. The only work done on internal friction at low temperatures is due to Bordoni.[15] Working in the kilocycle frequency range, he finds a peak in the

[11] See Table I, reference a.
[12] See Table I, reference c.
[13] R. R. Hasiguti and T. Hirai, J. Appl. Phys. **22**, 1084 (1951).
[14] A. H. Lawson, Phys. Rev. **60**, 330 (1941).
[15] P. G. Bordoni, Ricerca Sci. **19**, 851 (1949).

internal friction at about 1/3 the Debye θ of the metal. The height of the peak can be increased by cold working. In Table III is also given the temperature at Bordoni's peak for the metals he investigated. At lower temperatures the decrement approaches zero. These peaks perhaps are connected with the fact that they occur at a temperature when the period of vibration becomes comparable to the time required for a dislocation to break away from impurity atoms by thermal fluctuations. If they are due to a mechanism such as vacancy movement they would, of course, mask the effect due to dislocation movement.

At higher temperatures (above room temperature) the internal friction of metal single crystals is observed to increase with increasing temperature. This increase is usually gradual in the region of room temperature. For example, Nowick[5] found in a copper crystal an increase in the decrement from 2.5 to 4×10^{-4} in going from -49°C to 33°C. The increase in the decrement of lead[3] is also gradual up to 100°C. A qualitative explanation of this gradual increase in terms of the present theory is as follows: In the derivation of Eq. (2) the distribution of dislocation lengths which can break over potential hills of height σ to $\sigma+d\sigma$ that was used is $N(\sigma)d\sigma=Nd\sigma/\sigma_M$ for $\sigma<\sigma_M$ and $N(\sigma)d\sigma=0$ for $\sigma>\sigma_M$. Consider what effect thermal fluctuations will have on this distribution function. The effective height of each hill will be lowered because thermal fluctuations will aid the dislocation in getting over it. Hence the dislocations will all be shifted to lower values of σ in the distribution function. However, the hills of lowest σ will be effectively wiped out by the thermal fluctuations. Dislocations lying next to them will be restrained from large movements by adjacent higher hills. These dislocations are thus shifted to higher values of σ in the distribution function. $N(\sigma)d\sigma$ thus becomes at higher temperatures approximately equal to $Nd\sigma/\sigma_M{}^{\ddagger}$ where $\sigma_M{}^{\ddagger}$ is the reduced value of σ_M caused by thermal stress fluctuations. By Eq. (2) the internal friction is thus increased by the decrease in σ_M.

Another way for σ_M to decrease with increasing temperature is by the dissolving of any impurity atmospheres around the dislocations. An inspection of Table II shows that large changes in σ_M of amounts ranging from 10 to 1000 are possible by this means, which gives rise to large increases in the internal friction.

The only direct experimental evidence of the behavior of both σ_M and the internal friction as a function of temperature is for lead.[3] Up to about 200°C the decrease in σ_M is of a large enough order of magnitude to account for the increase in the decrement. However, at 300°C the change in the decrement is a factor of ten too large to be explained in this way. It is not too surprising that the present theory should prove inadequate to explain the internal friction at temperatures approaching the melting point of the metal. The complexity of the situation in the interior of the metal also increases with increasing temperature. Impurities are able to move rapidly about, the concentration of vacancies increases and dislocations are even able to move by climbing in directions other than slip directions. These changes in the metal would alter the conditions for which Eq. (2) was derived. The present theory should be considered only at temperatures well below the melting point of the metal.

TABLE III. Temperature below which the dislocations are effectively pinned down for various frequencies, assuming that $\epsilon=0.1$.

Metal	Temp. frequency 10^4 cps	Temp. frequency 10^6 cps	Temp. frequency 10^8 cps	Temperature at Bordoni peak
Pb	37°K	48°K	67°K	35°K
Ag	82	107	150	65
Cu	93	119	167	100
Al	103	132	186	95
Zn	132	170	237	

MODERATELY COLD-WORKED METALS

When a single crystal is cold-worked by stresses above the yield point there will be a multiplication of dislocations. As long as the increase in number is small so that the dislocations do not interact with each other, the effect of this increase will simply be to increase the internal friction and decrease the modulus. However, a point will be reached where they will interact with each other. The effect of this interaction, if the dislocations are in unstable positions just after cold working, is to move them to more stable positions. This movement would occur under a low temperature anneal. In equilibrium positions the dislocation interactions would cause the dislocations to feel a restoring force when displaced slightly. As dislocations move under a low temperature anneal to more stable positions the restoring force constant would increase. This increase would decrease the dislocation motion under an oscillatory stress and thus decrease the energy losses. This decrease in the internal friction of cold-worked materials under low temperature annealing is what Nowick calls the "Köster" effect.

Although the dislocations can move about under low temperature anneals their number is not expected to decrease much since, for example, the hardness of cold-worked copper does not decrease under low temperature annealing. If the energy lost per unit length of dislocation line remains unchanged after the specimen is cold-worked and room-temperature-annealed, the decrement ought to be higher than before cold working, since N, the length of dislocation lines, has increased. There have been several experiments which show that the internal friction, after an initial rise, can actually decrease under moderate cold work under conditions where the dislocations are not expected to anneal out of the crystal.[12–14,16] The reason for this has been under-

[16] See Table I, reference e.

stood for a long time.[14] Dislocation interaction suppresses dislocation motion and hence decreases the energy loss per unit dislocation length. To get a decrease of internal friction with increasing deformation it is only necessary that the energy lost per unit dislocation length decreases faster than the increase in dislocation length due to cold working. In high purity copper this phenomenon should occur. Weertman and Koehler[13] have made a rough calculation of the restoring force due to dislocation interaction in copper. If x is the displacement from the equilibrium position, the restoring force is approximately $3\mu b^2 Nx$ per unit length. When this restoring force, for a displacement λ, is much larger than $\sigma_M b$, the hysteresis phenomenon will be suppressed. For example, for a copper crystal with a value of $c=10^{-4}$ and $\epsilon=0.1$, N has to be greater than 3×10^8 in order to begin to suppress the hysteresis mechanism for oscillatory stresses less than σ_M. Now a value of N equal to 10^8 is that usually assumed for the dislocation density in a well-annealed specimen.[9] The dislocation density does not have to increase much to start suppressing the hysteresis mechanism. In high purity copper, stresses only slightly in excess of the yield point are sufficient to cause the internal friction to decrease with increasing cold work. In less pure materials greater amounts of cold work will be necessary to suppress the hysteresis. This explanation of the peak in the internal friction-cold work curve of copper we believe is more satisfactory than one previously presented.[12]†

When the hysteresis is completely suppressed the internal friction and modulus changes can be calculated from the equation of motion (see last section) of a dislocation in a potential well. For this case the decrement in the kilocycle frequency range is equal to

$$\text{decrement} \sim \frac{0.08\pi^2\mu b^2 N B\nu}{K^2} \sim \frac{0.08\pi^2\mu b^2 N B\nu}{(3\mu b^2 N)^2}, \quad (4)$$

where ν is the frequency of vibration, B is the dissipative force constant on a moving dislocation, and K is the restoring force constant. As an example of the magnitude of this term, if $N=3\times10^8/\text{cm}^2$ after cold working, $\nu=4\times10^4/\text{sec}$, and the theoretical value of B is used,[17–19] 5×10^{-5} dyne-sec/cm^2, this term is equal to 24×10^{-6} for lead. This value is quite a bit smaller than the internal friction as calculated from Eq. (2). The value of the internal friction given in Eq. (4) is proportional to the frequency, so an experimental determination of the frequency dependence would indicate whether the process given in Eq. (2) or Eq. (4) is the mechanism of internal friction in a moderately cold-worked material. It should be noted that the mechanism giving rise to Eq. (4) also suppresses any amplitude dependence of the internal friction. This suppression is experimentally observed.[12]

The suppression of the hysteresis mechanism through cold work will not bring about an increase in the modulus corresponding to the decrease in the decrement. This result can be shown with the aid of Eq. (7) of the next section. Equation (7) gives the fractional change in the modulus for the case when the hysteresis mechanism is unimportant. In the kilocycle range of frequency all terms in Eq. (7) containing the frequency are negligible. Dropping them out Eq. (7) reduces to $(\mu-\bar{\mu})/\mu=0.013$ when K is set equal to $3\mu b^2 N$. Since this value of the fractional change in the modulus is larger than those predicted by Eq. (3) for reasonable impurity concentrations, the modulus will continue to decrease as the hysteresis mechanism is suppressed even though the decrement is also decreasing. This behavior has been experimentally observed.[12]

MEGACYCLE FREQUENCY RANGE

The discussion so far has been concerned only with the internal friction as measured in the kilocycle or lower frequency range. For this region most of the experimental evidence indicates that the internal friction is independent of frequency. Figure 1 shows some recent work of Marx[20] on the internal friction of single crystals of copper as a function of frequency. The decrement is indeed almost independent of frequency. The data cover a wide enough range to eliminate any errors due to variations in the structure from one end of the crystal to the other. Nowick's hysteresis theory thus gives the right frequency dependence as well as a reasonable order of magnitude of the internal friction.

In the megacycle and higher frequency range a mechanism proposed by Koehler[2] should give an appreciable energy loss at low temperatures. Koehler had calculated the energy loss due to the motion of dislocations pinned down at points along their lengths by impurity atoms. He solved the equation of motion for the case where the frequency was in the kilocycle range. In Appendix II his equation of motion is solved for any frequency.

† In polycrystalline aluminum both Kê and Kunitomi [T. S. Kê, J. Metals **2**, 575 (1950); N. Kunitomi, Science Repts., Tôhoku Imp. Univ. **4**, 386 (1952)] have observed no decrease in the internal friction with increasing cold work. We feel that Kunitomi's explanation of this apparent discrepancy between the behavior of Al and Cu is probably correct.

In molybdenum Maringer [R. E. Maringer, J. Appl. Phys. **24**, 1525 (1953)] has shown that after an 8 percent strain with the load removed the decrement is about equal to the value before application of the load. This behavior is also found in copper. Since Maringer did not measure the decrement with the load removed at intermediate strains, it is impossible to compare his results with those of copper. The fact that the internal friction increased with increasing load (except for a drop at the yield point) without the load being removed is not inconsistent with the ideas advanced above. Under these conditions the effect of dislocation interaction is counterbalanced by the applied stress. On removal of the load the interaction is no longer counterbalanced and the internal friction should decrease. This behavior is what Maringer observed.

[17] J. D. Eshelby, Proc. Roy. Soc. (London) **197A**, 396 (1949).
[18] G. Leibfried, Z. Physik **127**, 344 (1950).
[19] F. R. N. Nabarro, Proc. Roy. Soc. (London) **209A**, 278 (1951).
[20] J. Marx, unpublished data.

At higher temperatures a model proposed by Eshelby[17] can be used. Eshelby assumed that a dislocation oscillates in a potential well. He wrote the equation of motion as

$$M\frac{d^2x}{dt^2}+B\frac{dx}{dt}+Kx=\sigma b\cos(2\pi\nu t). \tag{5}$$

Here M is the effective mass per unit length of the dislocation (M is equal to the density of the material times the square of the Burger's vector), K is the effective restoring force constant, and B is the damping constant.

Assuming that K is a constant the internal friction for a f.c.c. material is simply

$$\text{dec.}\sim\frac{0.08\pi^2\mu b^2 NB\nu}{(K-4\pi^2M\nu^2)^2+4\pi^2B^2\nu^2} \tag{6}$$

and the modulus change is given by

$$\frac{\mu-\bar{\mu}}{\mu}\sim\frac{0.04\mu Nb^2(K-4\pi^2M\nu^2)}{(K-4\pi^2M\nu^2)^2+4\pi^2B^2\nu^2}, \tag{7}$$

where $\bar{\mu}$ is the observed dynamic modulus and μ is the true modulus if no dislocations were present.

A characteristic value of K can be obtained from the average force constant a dislocation feels in its potential trough. Thus the average value of K is equal to $4\sigma_M b/\lambda$. Actually the dislocations will be distributed in troughs having values of K ranging from zero to the value $4\sigma_M b/\lambda$.

If we assume that all the dislocations feel the same applied stress and if we assume that the stress is so low that only a negligible number of dislocations are taking part in the hysteresis process, the total energy loss per cycle will be given by

$$\text{energy lost/cycle}=\frac{N\lambda}{4\sigma_M b}\int_{K=4\bar{\sigma}b/\lambda}^{K=4\sigma_M b/\lambda}\frac{2\pi^2B\nu\sigma^2b^2}{K^2+4\pi^2B^2\nu^2}dK$$

$$=\frac{\pi N\sigma^2b^2\lambda}{4\sigma_M b}\tan^{-1}\frac{K}{2\pi B\nu}\Bigg|_{K=4\bar{\sigma}b/\lambda}^{K=4\sigma_M b/\lambda}. \tag{8}$$

In this equation it is assumed that the mass of the dislocation can be neglected. The lower limit to the integral, $K=4\bar{\sigma}b/\lambda$, is set by the condition that only those dislocations which are not contributing an hysteresis loss are to be considered. At low frequencies, then, $\bar{\sigma}$ is simply equal to σ, the applied stress. At higher frequencies, $\bar{\sigma}$ may actually be smaller than σ since the period of vibration will become so small that it will be less than the time required for a dislocation to move over a potential hill.

When account is taken of the fact that the dislocations are distributed on many slip planes having

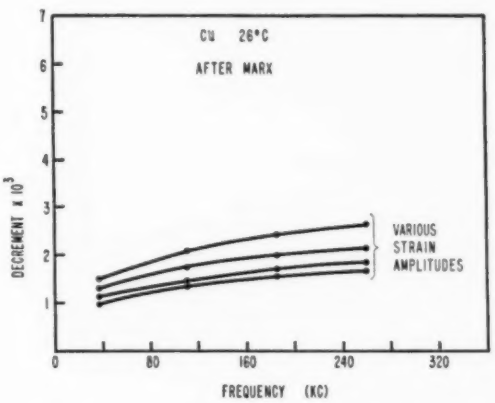


FIG. 1. The decrement of a 99.999 percent pure Cu crystal plotted as a function of frequency for various strain amplitudes. Data of J. Marx.

different values of σ, the decrement, using Eq. (8) is

$$\text{dec.}=\frac{0.01\pi N\mu b\lambda}{\sigma_M}\tan^{-1}\frac{K}{2\pi B\nu}\Bigg|_{K=4\bar{\sigma}b/\lambda}^{K=4\sigma_M b/\lambda}. \tag{9}$$

In the kilocycle range of frequency for a strain amplitude of vibration greater than 10^{-8} both $\tan^{-1}(2\sigma_M b/\pi\lambda B\nu)$ and $\tan^{-1}(2\bar{\sigma}b/\pi\lambda B\nu)$ are equal to $\pi/2$. The decrement by Eq. (9) is thus equal to zero. In the megacycle range for frequencies greater than $\nu^*=2\bar{\sigma}b/\pi\lambda B$ and less than $\nu^{**}=2\sigma_M b/\pi\lambda B$, the decrement by Eq. (9) is equal to

$$\text{dec.}=\frac{0.01\pi^2 N\mu b\lambda}{2\sigma_M}. \tag{10}$$

At frequencies greater than ν^{**} the decrement by Eq. (9) is again equal to zero.

Thus the decrement by Eq. (9) is zero at low frequencies, increases to the value given by Eq. (10) at the frequency ν^*, then remains constant up to the frequency ν^{**}, at which point it decreases to zero.

The decrement due to hysteresis given by Eq. (2) remains constant with frequency up to the frequency ν^*. At this frequency the hysteresis losses go to zero because the period of vibration is smaller than the time it takes for a dislocation to go over a potential hill. The total decrement which is the sum of those given by Eqs. (2) and (9), however, remains constant up to the frequency ν^{**} since the decrements by Eqs. (2) and (10) are approximately equal. If the oscillatory stress is equal to σ_M, the hysteresis phenomenon is active up to the frequency ν^{**}.

In extremely high purity materials (better than 99.999 percent) where the influence of the impurities becomes negligible the dislocations oscillate as loops of lengths equal to that of the Frank-Read sources. In the appendix it is shown that if all the loops are of length equal to l, the decrement reaches a maximum of magnitude equal to $0.04Nl^2$ at a frequency given by the

equation

$$\frac{2\pi\nu B}{2c}l^2=\frac{\pi^2\nu Bl^2}{3\mu b^2}=\frac{\pi^2}{4}.$$

The length of a Frank-Read source is usually assumed to be of the order of 10^{-4} cm. For this value the peak in the internal friction would occur at a frequency near 8×10^8 cps and would have a value equal to 0.01 if $N=10^8/\text{cm}^2$ and $B=5\times10^{-5}$ dyne-sec/cm.

In the megacycle frequency range another dissipative mechanism arising from the absorption and emission of vibrational quanta by conduction electrons has been considered by Paranjape.[21] He calculates that this mechanism will produce a frequency-independent decrement of order of magnitude equal to 10^{-3} for frequencies greater than 10^6 cps. If this mechanism be true the total energy loss would be the sum of it and the dislocation losses in this frequency range.

Another dissipative mechanism which is of importance in the megacycle range when longitudinal waves are used is that due to thermoelastic currents.[22] At toom temperature for ordinary metals a peak in the internal friction-frequency curve should occur near 10^{10} cps. The decrement at the maximum should be of the order of 10^{-3}.[22] In the liquid helium range of temperature the peak should occur at lower frequencies but its height should also decrease, going to zero at absolute zero.

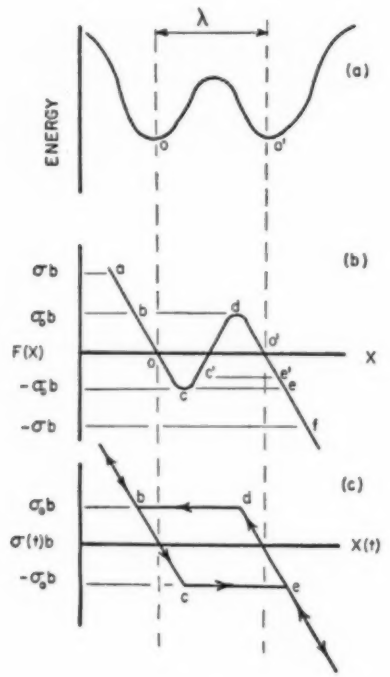


FIG. 2.

[21] B. V. Paranjape, Proc. Phys. Soc. (London) **66**, 572 (1953).
[22] C. Zener, *Elasticity and Anelasticity of Metals* (University of Chicago Press, Chicago, 1948), p. 90.

CONCLUSION

The hysteresis phenomenon of Nowick gives a reasonable order of magnitude for the experimentally observed internal friction and changes in Young's modulus for annealed and slightly cold-worked metals. An increase in the number of dislocations by a factor of 10 to 100 will start to suppress the hysteresis mechanism. This conclusion agrees with the experimental result in copper that the internal friction is reduced and strain amplitude dependence eliminated by stresses only slightly greater than the yield point. A frequency dependence measurement on cold-worked copper is needed to confirm this result.

Data on the internal friction in the megacycle range at low strain amplitudes on very pure metals should provide information for evaluating the damping constant of a moving dislocation.

ACKNOWLEDGMENTS

I want to thank Professor Nowick for allowing me to see his paper before publication and Dr. Marx for allowing me to reproduce Fig. 1.

APPENDIX I. CALCULATION OF THE ENERGY LOST DURING A HYSTERESIS CYCLE

Consider a unit length of a dislocation in a potential well such as is shown in Fig. 2a. For simplicity assume a symmetric well. When the dislocation is moved from the lowest energy position it will be acted on by a force. Call this force $F(x)$. Figure 2b indicates qualitatively what this force would be for the potential well given in Fig. 2a. Assume for the purpose of calculation that the unit length of dislocation line is the total length in a unit cube. If an oscillatory shear stress $\sigma(t)=\sigma\cos2\pi\nu t$ is applied to the unit cube, the dislocation will feel a force per unit length equal to $\sigma b\cos2\pi\nu t$. This force would set the dislocation into oscillatory motion. Let $x=x(t)$ be its position at any time t. The over-all strain s in the unit cube would be $s(t)=s_0(t)+s_1(t)$ where s_0 is the elastic strain and s_1 the strain due to the motion of the dislocation. Now the energy lost per cycle in the cube from this unit length would be equal to

$$\int_{\text{cycle}}\sigma(t)ds=\int_{\text{cycle}}\sigma(t)\frac{ds_1}{dt}dt$$

since

$$\int_{\text{cycle}}\sigma\frac{ds_0}{dt}dt=0.$$

The energy lost is simply equal to

$$\int_{\text{cycle}}\sigma(t)\frac{ds_1}{dt}dt=\int_{\text{cycle}}\sigma(t)b\frac{dx}{dt}dt=\int_{\text{cycle}}\sigma(t)bdx \quad \text{(i)}$$

since

$$\frac{ds_1}{dt}=b\frac{dx}{dt}.$$

In order to evaluate the right-hand integral of Eq. (i) it is necessary to know the equation of motion of the dislocation. We assume that it can be written as (neglecting the inertia of the dislocation)

$$B\frac{dx}{dt}-F(x)=\sigma(t)b, \qquad \text{(ii)}$$

where B is the dissipative force constant.

Consider now a very slowly varying cyclic stress applied to the dislocation whose magnitude σb is greater than $\sigma_0 b$. Assuming that the dislocation is at point **o** at time $t=0$, the dislocation moves from point **o** to **c** by the equation $\sigma(t)b=-F(x)$ since the term Bdx/dt can be made negligibly small by making ν small. At point **c** the dislocation is an unstable position. A small increase in the stress will cause it to move rapidly from **c** to **e**. The Bdx/dt term in Eq. (ii) now becomes important because $\sigma(t)b$ is no longer approximately equal to $-F(x)$. The speed at which the dislocation moves is given by Eq. (ii). Suppose at point **c'** $F(x)=-0.90\sigma_0 b$, then for $B=5\times10^{-5}$ [17–19] dynes-sec/cm^2, $\sigma_0=3\times10^3$ dynes/cm^2 (corresponding to a maximum strain in the unit cube of magnitude 10^{-8}) the speed at this point would be of the order of 6×10^{-2} cm/sec. The time required to go from **c'** to **e'** will be less than $\lambda/6\times10^{-2}$ sec or less than 10^{-5} to 10^{-6} sec. For larger values of σ_0 this time will be even smaller. Now if ν^{-1} is much larger than this time interval, the stress $\sigma(t)$ remains essentially constant when the dislocation moves from **c** to **e** or from **d** to **b**. Figure 2c shows what the hysteresis loop would be when the frequency is low enough for it to be considered "slowly varying." The energy loss per unit dislocation length is thus equal to $\sigma_0 b\lambda$. An unsymmetric potential well would not give precisely this answer but it would be of the same order of magnitude. Energy losses in the regions **a**–**b** and **e**–**f** are negligible compared to hysteresis losses in the kilocycle frequency range. This can be shown very simply. In these regions the energy loss is equal to

$$2\int_a^b B\frac{dx}{dt}dx+2\int_b^a B\frac{dx}{dt}dx\sim B\lambda^2\nu\sim2\times10^{-14}\ \text{erg/cycle}$$

for $\nu=10^4$/sec. Since $\sigma_0 b\lambda$ is of the order of 3×10^{-11} and becomes even larger for larger values of σ_0, it is the predominate term. In the megacycle range the frequency is no longer "slowly varying" and the damping term in Eq. (ii) can suppress the amplitude of motion to such an extent that the dislocation cannot move over the potential hill even if $\sigma>\sigma_0$.

APPENDIX II

Koehler[2] considered that the dislocation, under a small applied stress, took up the form pictured in Fig. 3. The coordinates x and y give the position of an element of the dislocation loop. The equation of motion

FIG. 3. Position of a "pinned down" dislocation under an applied stress.

of such an element is given by Koehler to be

$$M\frac{\partial^2x}{\partial t^2}+B\frac{\partial x}{\partial t}-C\frac{\partial^2x}{\partial y^2}=\sigma b\cos\omega t, \qquad \text{(a)}$$

where C is a constant equal to $3\mu b^2/\pi$ and $\omega=2\pi\nu$. By using an approximation method, Koehler solved this equation for the case where ν is in the kilocycle range. The steady-state solution for any frequency which satisfies the boundary condition that $x=0$ when $y=\pm l$ is

$$x=f\cos\omega t+g\sin\omega t, \qquad \text{(b)}$$

where

$$f=\left\{\frac{\sigma b}{(M^2\omega^4+B^2\omega^2)(\mathrm{Zin}^2\alpha,\beta l+\mathrm{Kas}^2\alpha,\beta l)}\right\}$$
$$\times\{(M\omega^2\,\mathrm{Kas}\alpha,\beta l+B\omega\,\mathrm{Zin}\alpha,\beta l)\,\mathrm{Kas}\alpha,\beta y$$
$$+(M\omega^2\,\mathrm{Zin}\alpha,\beta l-B\omega\,\mathrm{Kas}\alpha,\beta l)\,\mathrm{Zin}\alpha,\beta y\}$$
$$-\frac{\sigma bM\omega^2}{M^2\omega^4+B^2\omega^2},$$

$$g=\left\{\frac{\sigma b}{(M^2\omega^4+B^2\omega^2)(\mathrm{Zin}^2\alpha,\beta l+\mathrm{Kas}^2\alpha,\beta l)}\right\}$$
$$\times\{(M\omega^2\,\mathrm{Zin}\alpha,\beta l-B\omega\,\mathrm{Kas}\alpha,\beta l)\,\mathrm{Kas}\alpha,\beta y$$
$$-(M\omega^2\,\mathrm{Kas}\alpha,\beta l+B\omega\,\mathrm{Zin}\alpha,\beta l)\,\mathrm{Zin}\alpha,\beta y\}$$
$$+\frac{\sigma bB\omega}{M^2\omega^4+B^2\omega^2},$$

where

$$\mathrm{Kas}\alpha,\beta y=\cosh\alpha y\cos\beta y,$$
$$\mathrm{Zin}\alpha,\beta y=\sinh\alpha y\sin\beta y,$$
$$\alpha\beta=B\omega/2c,$$
$$\alpha^2-\beta^2=-M\omega^2/c.$$

For the case when $\alpha^2\approx\beta^2$ the decrement (assuming all loops have the same length and no dislocations are breaking away from impurity atoms) will be, for f.c.c. metals,

$$\text{decrement}=\frac{0.04\pi N\mu b^2}{B\omega}$$
$$\times\left\{1-\frac{\mathrm{Kas}\alpha,\alpha l(\sinh\alpha l\cos\alpha l+\cosh\alpha l\sin\alpha l)}{\alpha l(\mathrm{Zin}^2\alpha,\alpha l+\mathrm{Kas}^2\alpha,\alpha l)}\right.$$
$$\left.-\frac{\mathrm{Zin}\alpha,\alpha l(\cosh\alpha l\sin\alpha l-\sinh\alpha l\cos\alpha l)}{\alpha l(\mathrm{Zin}^2\alpha,\alpha l+\mathrm{Kas}^2\alpha,\alpha l}\right\}.$$

When $\alpha l \ll 1$ this expression reduces to Koehler's expression, which is proportional to the frequency. When $\alpha l \gg 1$, the decrement is inversely proportional to the frequency. When $\alpha l \sim \pi/2$ the internal friction reaches a maximum whose magnitude is given by

$$\text{dec.}_{\max} \sim \frac{0.04 N \mu b^2 l^2}{C} = 0.04 N l^2 \qquad \text{(c)}$$

if it is assumed that l is a constant. Actually l may have a range of values in a real crystal.

Note added in proof.—One very important contribution to the internal friction of cold-worked metals could arise from the vacancies and interstitials produced in cold working. Hasiguti [R. R. Hasiguti, J. Phys. Soc. Japan **8**, 798 (1953); *The Internal Friction of Metals due to Crystal Imperfections* (unpublished)] has proposed that movements of vacancy pairs, interstitial pairs, clusters of three or more vacancies, etc., will give rise to a contribution to the internal friction. Recently, Hasiguti, Hirai, and Kamoshita [Hasiguti, Hirai, and Kamoshita (to be published)] have observed a peak in the internal friction-temperature curves of cold-worked copper single crystals in the 22°–45° C region (kilocycle frequency range). These peaks decayed away a day or so after cold working. Hasiguti (private communication) has also several peaks in cold-worked iron in the region 0°–30°C. Since dislocations are not expected to give rise to a peak in this temperature range, Hasiguti's proposal seems reasonable. The occurrence of the higher temperature peak does not rule out the possibility that Bordoni's peaks are also caused by Hasiguti's mechanism, since one peak may be caused by vacancy pair movement and the other by interstitial pair movement.

Granato and Lücke [A. Granato and K. Lücke, "Internal friction and modulus changes due to dislocation damping in undeformed crystals," Technical Report, Brown University, Contract No. DA-36-039 SC-52623 (August 15, 1954)] have obtained a solution of Koehler's equation but in a different form than is given in this paper. Their solution has the advantage that damping losses are more easily calculable when the loop lengths are given by various distribution functions.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 2 FEBRUARY, 1955

Anomalous Forward Switching Transient in *p-n* Junction Diodes*

N. T. JONES, R. H. KINGSTON, AND S. F. NEUSTADTER
Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received January 25, 1954)

A delay in the flow of forward current when a grown-crystal *p-n* junction diode is switched from reverse to forward bias is explained on the basis of an extra *p-n* barrier in the grown-crystal bar. This effect was observed in 10 out of 24 production units, while no such anomaly was found in fused-junction diodes. A mathematical theory of the effect gives good agreement with the experimental results.

I. INTRODUCTION

DURING studies of the switching transients in grown-crystal germanium *p-n* junction diodes, one of the authors (N.T.J.) observed an anomalous behavior in several samples of early production models. When the diodes were placed in the circuit of Fig. 1, the current transient could be observed when the applied bias was switched from reverse to forward polarity. Figure 2 shows these transients on two different units: the curve shown in Fig. 2(a) is the anomalous case, and Fig. 2(b) shows the transient observed with a normal diode. As may be seen in the figure, the normal behavior shows an abrupt increase in current, followed by a gradual rise asymptotic to the steady-state value. In the anomalous case, the current is initially small (approximately 30 microamperes), and rises with increasing slope to an abrupt break, at which point it asymptotes to the final steady-state value. The normal switching transient has been reported previously by Waltz[1] and can be explained in a straightforward manner using the theory developed by Shockley.[2] Specifically, the initial current flow is given by the ratio of the applied voltage to the sum of the resistances of the *p*- and *n*-type sections of the diode, while the final slightly higher value of the current is caused by the decrease in these resistances because of minority carrier injection.

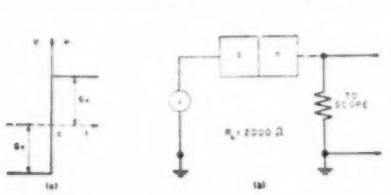


FIG. 1. Test circuit for observation of switching transient.

II. EXPLANATION OF EFFECT

Since it was impossible to explain the anomalous case on the basis of normal diode theory, even with a resort to traps, it was necessary to postulate one or more extra junction barriers somewhere in the structure. The very small initial current flow would otherwise indicate that practically all the voltage drop appeared across the true junction. The resultant ratio of injected minority carrier density to the normal den-

* The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.

[1] M. C. Waltz, Proc. Inst. Radio Engrs. **40**, 1483 (1952).
[2] W. Shockley, *Electrons and Holes in Semiconductors* (D. van Nostrand Company, Inc., New York, 1950), Ch. 12.

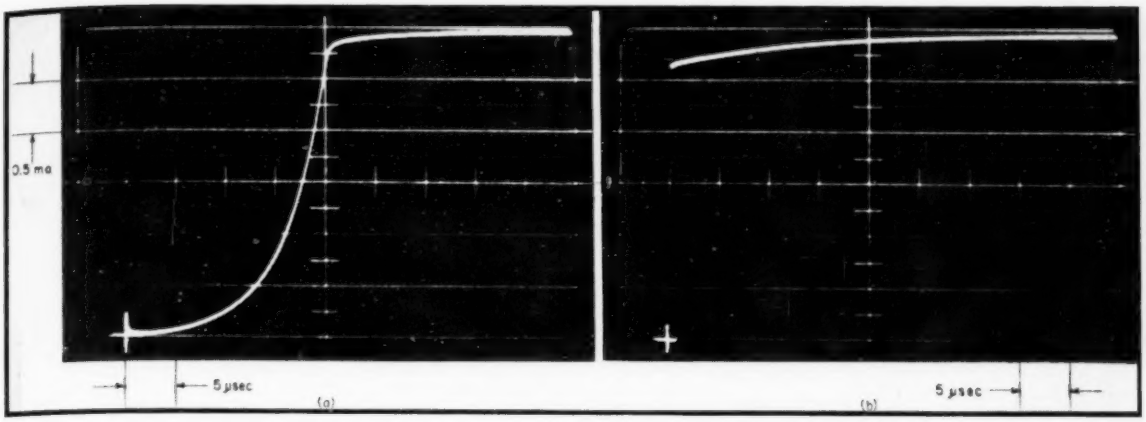


FIG. 2. Forward switching transient. (*a*) Anomalous case; (*b*) normal case. Both transients start at the + sign.

sity would then be[2]

$$\frac{p}{p_0}=\frac{n}{n_0}\approx e^{qV/kT}\approx e^{(39)6}=e^{234}, \tag{1}$$

which is not only completely unreasonable in itself, but also would lead to comparably large diffusion currents. All the diodes under study were known to have *n*-type resistivities in the range from 1 to 10 ohm-cm and *p*-type resistivities in the range from 0.01 to 0.1 ohm-cm. Because of the high resistivity of the *n*-type germanium, it was first postulated that the extra barrier occurred at the solder contact on this material. Such a contact might behave like the "formed" collector in a point contact transistor, thus giving the required barrier. To test this hypothesis, one of the units was removed from its plastic coating and probed with a tungsten point. The voltage transient between the probe point and the lead to the *n*-type material had only a small value when the point was near the end of the crystal bar adjacent to the *n*-type contact. By moving the point nearer to the true junction barrier, however, a large pulse, of the duration of the observed delay, was found to occur if the probe were closer than 0.2 mm to the *p*-type material. This, then, was the approximate position of the extra barrier, and from the mathematical treatment given below it was known that such a barrier must have an α, or current-multiplication factor, greater than unity. On this basis, the structure shown in Fig. 3 explains the anomalous transient, since the double barrier at the right of the first *n*-type zone, called a "hook" collector,[3] is known to have the necessary current multiplication. Evidently, because of inhomogeneities in the molten germanium when the single crystal was grown, the heavily doped *p*-type material was preceded by a small zone of *p*-type doping, since the standard growing procedure would start with the high-resistivity *n*-type end of the crystal.

In order to explain the general shape of the anomalous transient, it is necessary to look more closely at the mechanism of the buildup process. Figure 4 is a schematic representation of the structure of Fig. 3, where the region between the dashed lines represents the *n*-type material between the two *p*-type regions in Fig. 3. At the left barrier, or true junction, all current is assumed to be carried by holes, since the *p*-type material has a resistivity much less than that of the *n*-type material. This injected current i_{pl} is equal to the sum of the currents at the right barrier or "hook" collector. These three currents are: i_{n0}, the voltage-independent reverse current of the collector with no hole injection from the *n*-type region; i_{pr}, the hole current reaching the collector barrier; and i_n, the electron current caused by current multiplication at the collector. This representation of a "hook" collector follows that given by Shockley, Sparks, and Teal.[3] Now, if the diode is initially reverse-biased, the hole density in the central region will be small since W is less than the diffusion length for holes. Therefore, after switching to positive bias, the only current to flow will

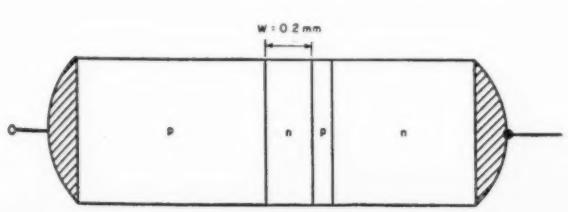


FIG. 3. Postulated structure of junction diode.

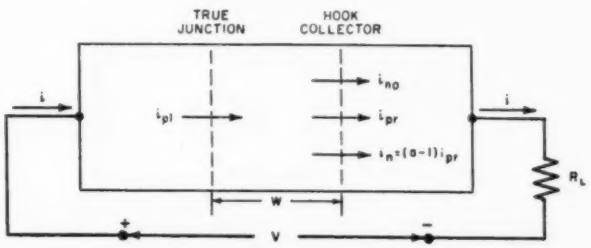


FIG. 4. Schematic representation of diode behavior.

[3] Shockley, Sparks, and Teal, Phys. Rev. **83**, 151 (1951).

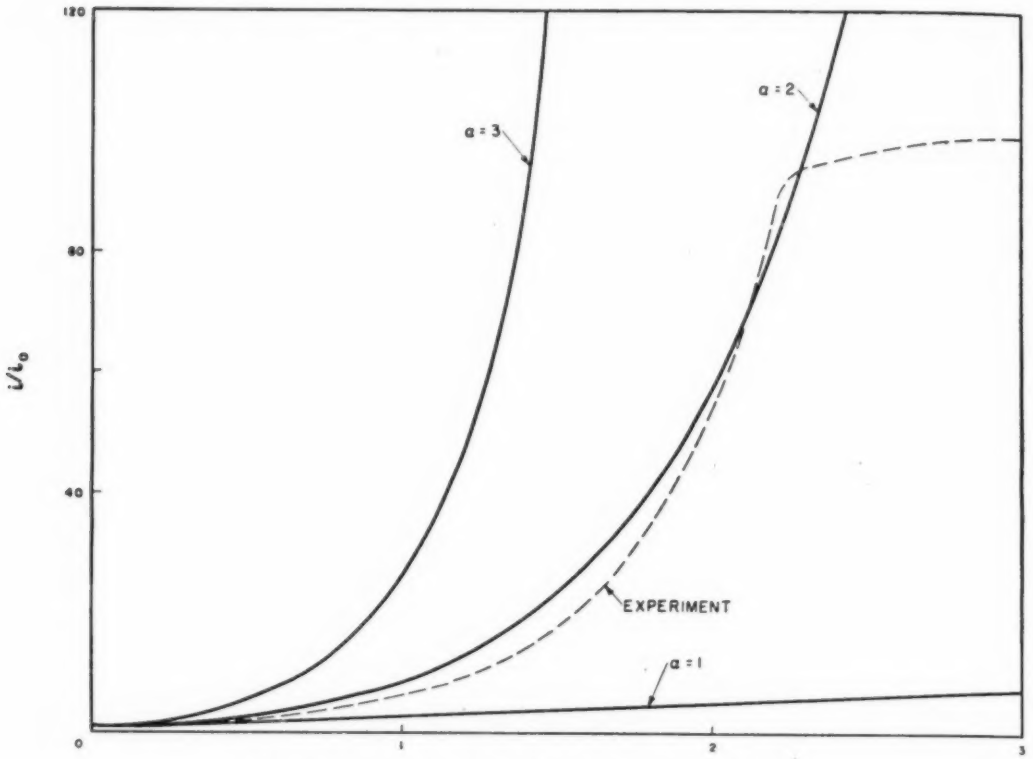


FIG. 5. Theoretical and experimental current transients.

be i_{n0}, and by continuity a hole current of magnitude i_{n0} will flow into the n-type material across the true junction. After a finite time, these holes will reach the collector, causing an increase in i_{pr} and a consequent additional current i_n. It may be seen that this process is regenerative in the sense that any increase in the current at the collector must be accompanied by an equal increase in the hole current injected at the left barrier. The mathematical treatment of this problem is given in the appendix, and involves the continuity relationship

$$i=i_{pl}=i_{n0}+i_{pr}+i_{pr}(\alpha-1), \qquad (2)$$

and the time dependence of i_{pr} with a given variation of i_{pl}. The latter relationship is a critical function of the width of the region W and the diffusion constant for holes D_p. The results of the mathematics are plotted in Fig. 5 which shows the ratio of instantaneous to initial current as a function of the reduced time parameter $D_p t/W^2$. The collector will behave as a barrier and current multiplier only so long as the voltage across it is much greater than kT/q; therefore, when the current in the case discussed reaches a value near 3 milliamperes, this voltage becomes small and the barrier effectively disappears. This explains the abrupt break in the buildup of the forward transient and the final gradual approach to equilibrium caused by attainment of constant injected hole density in the n-type material. With this in mind, the experimental curve in Fig. 5 shows good agreement with the theory and, using the value of W (0.2 mm), gives an α for the collector of approximately 2. This value of α may be shown to be quite consistent with the reverse current of the hook i_{n0}. A further verification of the theory was obtained in two other qualitative tests. If the model is correct, increasing the hole concentration in the central region prior to switching should produce a much higher initial current and consequently a much more rapid buildup. This effect was observed in two ways. First, if the diode was initially unbiased, then the anomalous effect was never observed. This means that the equilibrium hole density in the n-type region was high enough to cause an initial collector current flow comparable to 3 milliamperes, or the final steady-state value. Second, if an unpackaged diode was exposed to a concentrated light source, the delay time was observed to decrease and practically disappear as the light became brighter. The initial value of current increases accordingly. This again verifies the theory that added minority carriers raise the initial value of the forward current.

In all the diodes studied the dc voltage-current characteristics were normal except for a slightly higher forward resistance in those exhibiting the current delay. This is in agreement with the physical model.

III. DISCUSSION

The quantitative results described above were obtained on one particular sample; however, out of 24 diodes studied, 10 exhibited the anomalous effect to

varying degrees. For a range of n-type resistivities from 0.3 to 10 ohm-cm, the high-resistivity units were most prone to the phenomenon. This would seem to verify the picture of nonuniform doping, since the higher resistivity material could more easily be doped to the opposite conductivity type. The model of a single extraneous p-region may, of course, be an oversimplification of the structure since, in reality, there may be a much more complicated distribution of impurities. The above treatment certainly explains the general effect in spite of this fact. Further data have been obtained on the variation of the delay time with reverse voltage and final forward current. The qualitative results are in good agreement with the theory. It is significant that this anomalous effect could not be observed in fused-junction diodes,[4] since the fusion process would not be expected to give such an inhomogeneous distribution of impurities. In conclusion, it seems that in the selection of diodes for switching purposes, this possible anomalous behavior should be kept in mind and that appropriate tests should be performed.

ACKNOWLEDGMENT

The authors are indebted to J. E. Thomas, Jr., of this laboratory for encouragement and criticism of the work.

APPENDIX

Mathematical Treatment of Current Transient

Rewriting Eq. (2) of the text,

$$i = i_{n0} + \alpha i_{pr}, \tag{3}$$

and taking the Laplace transformation,

$$\bar{i} = \frac{i_{n0}}{s} + \alpha \bar{i}_{pr}. \tag{4}$$

If $i_0(t)$ is the current $i_{pr}(t)$ resulting from a normalized spike impulse $i(t) = \delta(t)$, then

$$i_{pr}(t) = \int_0^t i(t-\tau) i_0(\tau) d\tau. \tag{5}$$

[4] R. N. Hall, Proc. Inst. Radio Engrs. **40**, 1512 (1952).

Rewriting the minority carrier diffusion equation[2] in reduced form gives

$$\frac{\partial p}{\partial T} = \frac{\partial^2 p}{\partial X^2}, \tag{6}$$

where the lifetime has been assumed infinite, $T = D_p t / W^2$, and $X = x/W$.

Transforming Eq. (6) gives

$$s\tilde{p} = \frac{\partial^2 \tilde{p}}{\partial X^2}, \tag{7}$$

which must now be solved subject to the boundary conditions

$$-D \left.\frac{\partial p}{\partial x}\right|_{x=0} = \delta(t) \quad \text{or} \quad \left.\frac{\partial \tilde{p}}{\partial X}\right|_{X=0} = -1 \tag{8}$$

and

$$p(X=1) = 0, \tag{9}$$

since the hole density at the collector must always be zero. The resulting solution gives

$$\bar{i}_0(T) = \left.\frac{\partial \tilde{p}}{\partial X}\right|_{X=1} = \frac{1}{\cosh\sqrt{s}}, \tag{10}$$

and combining Eq. (10) with Eqs. (5) and (4),

$$\bar{i} = \frac{i_{n0}}{s} + \frac{\alpha \bar{i}}{\cosh\sqrt{s}} \tag{11}$$

by use of the convolution theorem.

Hence

$$\bar{i} = \frac{i_{n0} \cosh\sqrt{s}}{s[\cosh\sqrt{s} - \alpha]}, \tag{12}$$

and inverting,

$$\frac{i}{i_{n0}} = \frac{1}{1-\alpha} + \sum_{k=-\infty}^{+\infty} \frac{2\alpha \exp[\cosh^{-1}\alpha + 2k\pi i]^2 T}{(\alpha^2-1)^{\frac{1}{2}}[\cosh^{-1}\alpha + 2k\pi i]} \tag{13}$$

for $\alpha \neq 1$, and for $\alpha = 1$,

$$\frac{i}{i_{n0}} = 1 + 2T + 8T \sum_{k=1}^{\infty} \exp[-4\pi^2 k^2 T]. \tag{14}$$

Methods of Measuring Dielectric Constants Based upon a Microwave Network Viewpoint

A. A. Oliner and H. M. Altschuler
Microwave Research Institute, Polytechnic Institute of Brooklyn, Brooklyn, New York
(Received November 25, 1953)

Dielectric samples located in wave guide are represented by network four-poles whose admittance determinant is related simply to the dielectric constant. The four-pole viewpoint, which has heretofore not been adequately exploited, makes possible the utilization of the whole body of transmission line four-pole measurement techniques, including those of precision and invariance. New measurement procedures based on these techniques have been formulated which are either length invariant or location invariant; that is, the dielectric constant is determinable without the knowledge of either the length or the location of the sample. Four specific measurement procedures are presented which incorporate the precision as well as the invariance aspects.

I. INTRODUCTION

THE viewpoint which regards the dielectric sample located in wave guide as a four-pole is not in itself new. Its validity has been recognized independently by a number of writers, including Westphal[1] and one of the present co-authors,[2] but its implications have heretofore not been fuliy exploited. The measurement procedures of this type listed by Westphal simply relate the dielectric constant to the input impedances corresponding to specific output terminations. Perhaps the most well-known of these procedures employs an open and a short circuit termination; this method was published independently by first British,[3] and then American,[4] authors. These methods, based on specific output terminations, utilize the four-pole viewpoint in only a relatively narrow sense since no attempt has been made to incorporate any of the more highly developed aspects of four-pole measurement techniques.

The measurement methods presented here, which are based in part on material in reference 2, relate the dielectric constant to the determinant of the admittance representation of the dielectric sample, so that the measurement procedure is reduced to finding the parameters of an equivalent network characterizing it. Of the body of four-pole measurement techniques available in this context, particular use is made herein of certain precision measurement methods and of the principle of invariance.

Although the need for increased precision in the determination of the network parameters of microwave structures has of late given rise to a number of precision procedures, in contrast to point procedures,[5] transmission line methods of measuring dielectric constant have been confined to point measurements, as far as the authors are aware. It is sufficient here to indicate the most appropriate of the precision methods which can be taken over bodily from four-pole measurements into dielectric measurements. When the dielectric constant is assumed to be purely real, or when tan δ is so small that ϵ' can be measured independently, the four-pole may be regarded as lossless and may be determined by the tangent relation method.[6] When the four-pole, i.e., the dielectric sample, is dissipative, one can employ the geometrical method in the reflection coefficient plane given by Deschamps.[7]

The concept of invariance, as used here, is related to the error due to physical length measurements. Since such errors are generally greater than any resulting from electrical measurements, it is desirable from the standpoint of increased accuracy to minimize the number of physical length measurements necessary. Other errors evidently exist in the measurement of dielectric constants which may be more significant in most cases than the length and location errors discussed here. These other errors are assumed to be corrected for in standard fashion.[8] Invariance is treated in detail in Sec. II. 2, where considerations are presented which permit a measurement of dielectric constant independent either of the length of the sample or of its location in the wave guide. Although the four-pole type measurement procedures listed by Westphal[1] do not directly involve

[1] W. B. Westphal *et al.*, "Techniques of Measuring the Permittivity and Permeability of Liquids and Solids in the Frequency Range 3 c/s to 50 kmc/s," Technical Report XXXVI, Laboratory for Insulation Research, M.I.T., July, 1950, Table 6.

[2] A. A. Oliner, "The Measurement of Dielectric Constants by the Use of the Tangent Relation Method," Laboratory Memorandum, Microwave Research Institute, Polytechnic Institute of Brooklyn, March, 1949.

[3] J. B. Birks, Proc. Phys. Soc. (London) **60**, 282 (1948).

[4] W. H. Surber and G. E. Crouch, J. Appl. Phys. **19**, 1130 (1948).

[5] By point measurements one means that only a specific, minimum number of datum points are taken, while precision measurements refer to measurements in which a sufficiently large number of different datum points are taken to be analyzed later by a technique which effectively averages them. Repetition of the measurement of one or more points is not, of course, included in the latter category.

[6] N. Marcuvitz, *Waveguide Handbook* (McGraw-Hill Book Company, Inc., New York, 1951), Radiation Laboratory Series **10**, Chap. III.

[7] G. A. Deschamps, J. Appl. Phys. **24**, 1046 (1953).

[8] The measured data is corrected for errors in the location of the voltage minimum due to spurious discontinuities; errors in VSWR due to detector nonlinearity; and errors in VSWR due to dissipation in the length of guide in front of the sample. From the data so corrected the network representing the sample is derived and its parameters are then corrected for errors introduced by dissipation in the guide behind the sample and in the short circuit. The complex dielectric constant obtained from the network is finally corrected for errors caused by wall loss and imperfect fit in the guide surrounding the sample.

the sample length, the location of both faces of the sample must be determined via two physical length measurements, only one of which is generally trivial. The length invariant procedures given here require the location of only one of the sample faces, a step which can easily be made trivial. For a location invariant measurement one need know only the length of the sample.

Depending on the circumstances, a length invariant or a location invariant measurement may be more suitable. For example, in rectangular wave guide where one face of the dielectric sample can be located accurately, a length invariant measurement is indicated. In coaxial line, however, where the frequencies are generally lower and where the terminal planes are not as easily determined, a location invariant procedure may often be preferable. In Sec. III, several explicit measurement procedures are presented which are not intended to be inclusive but which cover a number of practical situations.

The averaging and extrapolation schemes employed in four-pole measurements may also be used to advantage in such point measurement procedures as that of Roberts and von Hippel[9] or Birks[3] and Surber and Crouch.[4] These averaging procedures and their applications to point measurements are treated in Sec. II. 3.

II. UNDERLYING PRINCIPLES

1. Dielectric Constant Derived from Network Representations of Sample

A dielectric sample in wave guide is regarded as a four-terminal structure at the terminal planes $T_{\epsilon 1}$ and $T_{\epsilon 2}$ defined by the air-dielectric interface as shown in Fig. 1(a). In view of its symmetry, this structure may be represented, for example, by the symmetric *Pi* network shown in Fig. 1(b). Since there are no discontinuities at $T_{\epsilon 1}$ and $T_{\epsilon 2}$ apart from the change in characteristic impedance, a relation between the properties of the sample and the parameters of the network follows readily upon bisecting both the structure and the network of Fig. 1, and successively placing an open and a short circuit at the mid-planes. One then obtains

$$\begin{aligned} Y_{11}-Y_{12}&=jY_\epsilon \tan(\kappa_\epsilon L_\epsilon/2),\\ Y_{11}+Y_{12}&=-jY_\epsilon \cot(\kappa_\epsilon L_\epsilon/2) \end{aligned} \tag{1}$$

upon equating the resulting input admittances. The product of these equations divided by Y_0 yields

$$(Y_\epsilon')^2=(Y_{11}')^2-(Y_{12}')^2\equiv \mathcal{Y}_\epsilon. \tag{2}$$

In the above equations, the prime denotes normalization to Y_0, the characteristic admittance of the air filled guides; Y_ϵ and κ_ϵ are, respectively, the characteristic admittance and the propagation wave number of the dielectric filled guide, and $\mathcal{Y}_\epsilon$, as defined, is the admittance determinant, normalized to Y_0, of the symmetric representation at the planes $T_{\epsilon 1}$ and $T_{\epsilon 2}$.

From the well-known dependence of Y_ϵ' on the relative dielectric constant ϵ_r, one finds at once that

$$\epsilon_r=\mathcal{Y}_\epsilon \tag{3a}$$

for TEM modes (dominant mode in coaxial line), and

$$\epsilon_r=\frac{\mathcal{Y}_\epsilon+(\lambda_g/\lambda_c)^2}{1+(\lambda_g/\lambda_c)^2} \tag{3b}$$

for H modes (dominant mode in rectangular guide), where $\epsilon_r=\epsilon'(1-j\tan\delta)$. The guide wavelength and the cut-off wavelength (both in the air filled guide) are designated respectively by λ_g and λ_c. In view of the simple dependence of ϵ_r on $\mathcal{Y}_\epsilon$, any representation of the sample at $T_{\epsilon 1}$ and $T_{\epsilon 2}$ which is readily translatable into the admittance determinant $\mathcal{Y}_\epsilon$, will yield ϵ_r almost at once. For example, if the sample is represented at $T_{\epsilon 1}$ and $T_{\epsilon 2}$ in terms of normalized impedance (Z') or scattering (S) parameters, $\mathcal{Y}_\epsilon$ is related to these parameters by

$$\mathcal{Y}_\epsilon=\frac{1}{(Z_{11}')^2-(Z_{12}')^2}=\frac{(1-S_{11})^2-S_{12}^2}{(1+S_{11})^2-S_{12}^2}. \tag{4}$$

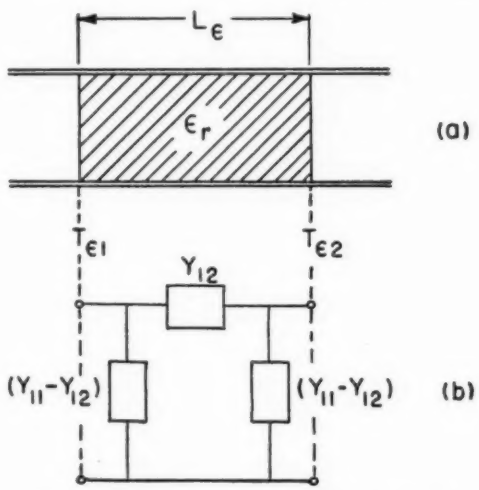


FIG. 1. Dielectric sample in a wave guide and one of its representations.

2. Location and Length Invariant Measurements

In equivalent circuit measurements, it has been recognized[10] that the geometrical measurement necessary to locate the terminal planes of the structure generally introduces a source of error greater than that resulting from any of the electrical measurements. Special representations,[11] particularly those involving

[9] S. Roberts and A. von Hippel, J. Appl. Phys. **17**, 610 (1946).

[10] N. Marcuvitz, reference 6, pp. 137–8.

[11] A. A. Oliner, "Preferred Equivalent Network Representations of Microwave Structures," Symposium on Modern Network Synthesis, pp. 236–246, Polytechnic Institute of Brooklyn, April, 1952.

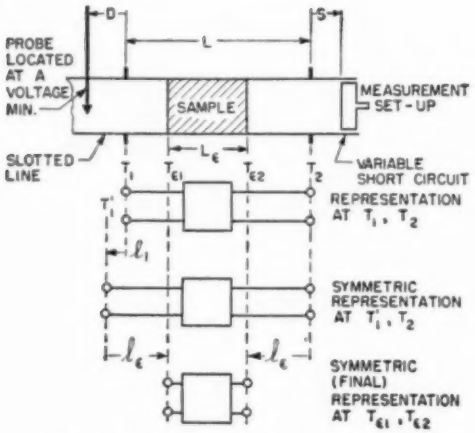


FIG. 2. Terminal plane shifts to obtain an invariant representation.

symmetry, may then be devised for which the parameters are determined independent of any physical length measurements, and therefore depend only on electrical measurements. These representations have been called "distance invariant."[10] The usual transmission line methods of measuring dielectric constant require the knowledge of the length of the sample and/or the location of the air-dielectric interfaces, both of which must be determined by a physical length measurement of some sort. Procedures for finding the dielectric constant which eliminate or minimize the number of these error sources, i.e., the physical length measurements, are therefore inherently capable of higher accuracy. The present section discusses the application of such invariance procedures to dielectric constant measurements.

Assume that a dielectric sample located arbitrarily in a wave guide at $T_{\epsilon 1}$, $T_{\epsilon 2}$ (see Fig. 2) has been measured as a four-pole with respect to two arbitrary reference planes T_1, T_2 and that a representation at these planes has been found. One can then obtain a symmetric representation at T_1', T_2 by shifting the input terminal plane a distance l_1 from T_1 to T_1'. The distance l_1 and the parameters of the symmetric representation at the new planes are both consequences of the symmetry demand only. We then seek the representation at the planes $T_{\epsilon 1}$ and $T_{\epsilon 2}$, from which the dielectric constant is readily found. In achieving this end, advantage is taken of the known symmetry of the physical structure at these planes by demanding a symmetric representation coupled with one of two additional stipulations.

One of these additional stipulations is that one of the two terminal planes of the final representation sought must coincide with $T_{\epsilon 2}$. In this case, as a consequence of shifting both terminal planes through a distance l_ϵ, the other terminal plane will automatically fall at $T_{\epsilon 1}$. One therefore has the desired representation without the knowledge of the sample length L_ϵ and without the need for locating the second terminal plane $T_{\epsilon 1}$. If, in this length invariant measurement, the sample has been located in the guide in such a manner that T_1 and $T_{\epsilon 1}$ coincide, no physical length measurement whatsoever is necessary. It is possible to do this in many practical situations to a high degree of accuracy.

Alternatively, one can make the additional stipulation that the terminal planes of the final representation being sought must be a distance L_ϵ apart. This condition, coupled with the symmetry restriction, will again yield the desired representation at $T_{\epsilon 1}$ and $T_{\epsilon 2}$ when one shifts both terminal planes by l_ϵ. In this case the location of the sample itself need not be known; one requires only the knowledge of the sample length L_ϵ and L, the distance separating the planes T_1 and T_2 with respect to which the measurement was originally carried out. It is therefore called a location invariant measurement. Furthermore T_1 and T_2 can usually be taken at one and the same physical plane so that L is identically zero.

3. Extension of the Measurement Range

The input reflection coefficient locus (at T_1) corresponding to all reactive terminations (at T_2) is known to be a circle, since the dielectric sample may be viewed as a four-pole. This fact can be employed to form the basis for averaging data and for extending somewhat the measurement range of the equipment when it is used in connection with techniques[1,3,4,9] originally conceived as point measurements.[5] More specifically, a number of input datum points, corresponding to various short circuit positions S, are obtained and plotted in the reflection coefficient plane. In addition, the positions of the input voltage minima D are plotted against S. The best curves fitting these data can then be drawn to yield, respectively, a circle and a D *vs* S curve, as illustrated in Fig. 3. For increased accuracy it is necessary to plot the relevant portions of the D *vs* S curve to a large scale. The desired information, namely, the input reflection coefficients corresponding to specified values of S, can be read from the two curves.[12] For instance, the input data required for the open circuit-short circuit measurement[3,4] can be obtained in this fashion by reading the values corresponding to $S=0$ and $S=\lambda_g/4$ from the curves.

When one portion of the input reflection coefficient locus falls into a region in which the associated values of VSWR are low enough to be measurable, while the rest of the locus falls outside this region (i.e., values of VSWR are too high to permit sufficiently accurate measurement), the known circular shape of the locus enables one to find any reflection coefficient lying outside the measurable range. One simply completes the circle defined by the measured points and determines whatever point may be desired via the two curves. If, on the other hand, *all* datum points fall outside the

[12] For a given value of S, the corresponding value of D is located on the D *vs* S curve. This value of D can be interpreted as θ, an angle in the reflection coefficient plane, so that the magnitude associated with this angle can be read from the circle.

measurable range, or when only a small number of points fall within this range so that a circle cannot be defined accurately, the shunt technique[13] may be employed as an indirect means of obtaining the high values of VSWR. Such schemes permit the use of specific point procedures for the measurement of samples of relatively low loss and arbitrary length. For example, when one employs the Roberts-von Hippel method[9] in its original sense, a quarter wavelength sample is recommended as a way of minimizing the value of VSWR to be measured. The use of the circle as a means of extrapolation, however, more readily allows the measurement of other than quarter wavelength samples without any change in the fundamental analysis. Similarly, in the open circuit-short circuit measurement, the open circuit point, say, may fall within and the short circuit point outside the measurable range. The use of the input reflection coefficient locus will then assure that both of these points can be found if a sufficiently large portion of the locus falls into the measurable range so that a circle can be completed.

III. SPECIFIC MEASUREMENT PROCEDURES

1. Dissipative Sample: Length Invariant Measurement

In the experimental setup shown in Fig. 4, the location of one (and only one) of the planes of the dielectric sample must be known accurately since the measurement is to be length invariant. The sample has consequently been placed exactly at T, the end of the guide containing it. All measurements are made with respect to the plane T; i.e., D, the position of the probe at a voltage minimum and S, the position of the short circuit, are both referred to T as shown. Initially, let a scattering representation S_{11}', S_{12}', S_{22}', be obtained at this plane by Deschamps' method.[7] Although T_- and T_+ are both identically T, the plane T_- can be thought

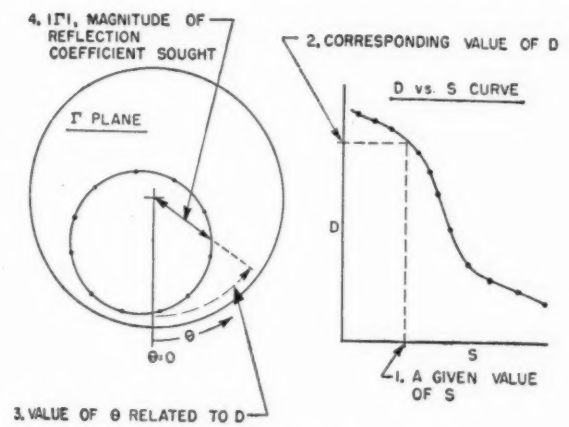


FIG. 3. Data curves.

[13] H. M. Altschuler and A. A. Oliner, "A New Shunt Technique for Microwave Measurements," Report R-326-53, PIB-262, Polytechnic Institute of Brooklyn, Microwave Research Institute, June, 1953.

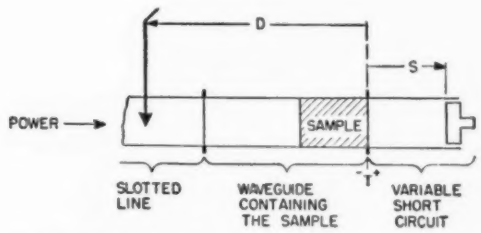


FIG. 4. Setup for a length invariant measurement.

of as being located $n\lambda_g$ to the left of T_+, where λ_g is taken in the air filled guide. The parameters can be expressed in the form:

$$S_{11}'=|S_{11}'|e^{j\sigma_{11}},\quad S_{12}'=|S_{12}'|e^{j\sigma_{12}},\quad S_{22}'=|S_{22}'|e^{j\sigma_{22}}, \tag{5}$$

where $|S_{11}'|=|S_{22}'|$, since the physical structure is symmetric. Let a symmetric representation S_{11}, S_{12}, S_{22} now be demanded by shifting the input plane T_- through a distance l_1 to T_1'. The condition for a symmetric representation is $S_{11}=S_{22}$, where

$$S_{11}=S_{11}'e^{-j4\pi l_1/\lambda_g}\quad\text{and}\quad S_{22}=S_{22}', \tag{6}$$

so that we have

$$S_{11}=|S_{11}'|e^{j(\sigma_{11}-4\pi l_1/\lambda_g)}=|S_{11}'|e^{j\sigma_{22}}=S_{22}', \tag{7}$$

from which one obtains

$$l_1=\frac{\lambda_g}{4\pi}(\sigma_{11}-\sigma_{22}). \tag{8}$$

Since the output plane is still located at T (which coincides with $T_{\epsilon 2}$), the input plane T_1' must now lie at $T_{\epsilon 1}$ in view of the symmetry. The mutual element at shifted planes is given by

$$S_{12}=S_{12}'e^{-j2\pi l_1/\lambda_g}, \tag{9}$$

where l_1 has been solved for above. The desired representation at $T_{\epsilon 1}$ and $T_{\epsilon 2}$ in terms of the measured quantities is therefore given by

$$S_{11}=|S_{11}'|e^{j\sigma_{22}};\quad S_{12}=|S_{12}'|e^{j[\sigma_{12}+(\sigma_{22}-\sigma_{11})/2]}. \tag{10}$$

The dielectric constant is computed from these relations via Eqs. (4) and (3).

It is to be noted that the dielectric constant has been determined without the knowledge of either the sample length or the location of the other air-dielectric interface. This is pointed out in contrast to other measurement procedures (e.g., references 3 and 4) in which the sample length does not appear but for which *both* sample faces must be located.

2. Dissipative Sample: Location Invariant Measurement

The experimental setup here is identical to that shown in Fig. 4, except that the sample, in this instance

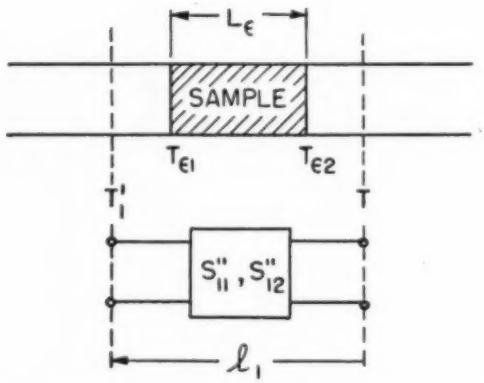


FIG. 5. Sample location and one of its representations for a location invariant measurement.

of known length L_ϵ, is located arbitrarily, i.e., T does not coincide with $T_{\epsilon 2}$ (see Fig. 5). The measurements, however, are carried out as before with respect to T so that the representation described by Eq. (5) is again obtained. All the steps leading up to Eq. (10) are followed here as well, except that now T_1' and T do not (yet) coincide with $T_{\epsilon 1}$ and $T_{\epsilon 2}$. At this point in the analysis the representation is as depicted in Fig. 5, where the parameters in Eq. (10) have been renamed S_{11}'', S_{12}''. In order to effect the shift to the planes $T_{\epsilon 1}$ and $T_{\epsilon 2}$ each reference plane must be moved through $-(l_1-L_\epsilon)/2$ to yield S_{11}, S_{12}, the symmetric representation with terminal planes L_ϵ apart; the parameters are

$$S_{11}=S_{11}''e^{-j2\pi(L_\epsilon-l_1)/\lambda_g},\quad S_{12}=S_{12}''e^{-j2\pi(L_\epsilon-l_1)/\lambda_g}, \tag{11}$$

where l_1 is given by Eq. (8). In terms of the parameters at T, obtained directly from the measurement, one has

$$\begin{aligned} S_{11}&=\pm|S_{11}'|e^{j[(\sigma_{11}+\sigma_{22})/2-2\pi L_\epsilon/\lambda_g]},\\ S_{12}&=|S_{12}'|e^{j[\sigma_{12}-2\pi L_\epsilon/\lambda_g]}. \end{aligned} \tag{12}$$

The use of Eqs. (4) and (3) again yields ϵ_r. The ambiguity in the sign of S_{11} is readily resolved.[14]

3. Lossless Sample: Length Invariant Measurement

The measurement is carried out here with respect to two distinct planes T_1 and T_2, with $T_{\epsilon 2}$, one of the faces of the sample, coincident with T_2 as shown in Fig. 6. Let the data be analyzed by the tangent relation method[6] to yield the parameters D_0, S_0, γ; the corresponding representation at T_1 and T_2 is indicated in Fig. 6.

One now needs to demand only that the input plane be shifted to yield a representation at $T_{\epsilon 1}$ and $T_{\epsilon 2}$. The values of γ and S_0 remain unchanged by the shift, while D_0 takes on the new value D_0'. However, D_0' need never be determined since it is known that, for a symmetric representation,[6]

$$\tan(2\pi S_0/\lambda_g)=-\cot(2\pi D_0'/\lambda_g). \tag{13}$$

It can be shown[6] that the admittance determinant in terms of D_0, S_0, and γ is expressed by

$$\mathcal{Y}_\epsilon=\frac{\tan(2\pi S_0/\lambda_g)-\gamma\tan(2\pi D_0'/\lambda_g)}{\tan(2\pi D_0'/\lambda_g)-\gamma\tan(2\pi S_0/\lambda_g)}, \tag{14}$$

so that, in view of Eq. (13),

$$\mathcal{Y}_\epsilon=-\frac{\tan^2(2\pi S_0/\lambda_g)+\gamma}{1+\gamma\tan^2(2\pi S_0/\lambda_g)}. \tag{15}$$

It is worth noting that this last result has been derived without any knowledge of the location of the planes T_1 and $T_{\epsilon 1}$, and that it is completely independent of the input terminal plane shift which yielded D_0'. In consequence, the locations of T_1 and $T_{\epsilon 1}$ need never be determined. The dielectric constant may be obtained from Eqs. (15) and (3) as soon as S_0 and γ (of Fig. 6) have been found.

If, in Fig. 6, the sample face $T_{\epsilon 2}$ were a distance L_2 removed from T_2, Eqs. (13) to (15) would still apply with S_0 replaced by S_0', where

$$S_0'=S_0+L_2. \tag{16}$$

4. Lossless Sample: Location Invariant Measurement

Let the measurement setup shown in Fig. 6 again be taken, except that in this case the sample is located arbitrarily, and that L_ϵ, the length of the sample, and L, the distance separating T_1 and T_2, are assumed to be known. Again, the representation at T_1 and T_2 shown in Fig. 6 and based on the numbers D_0, S_0, and γ is

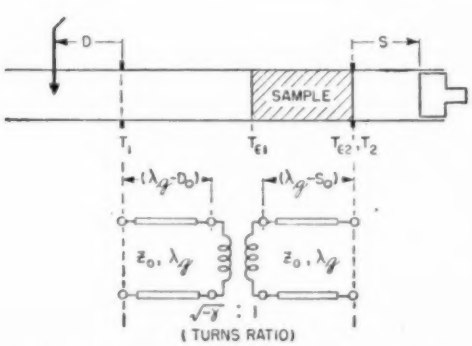


FIG. 6. Setup for a length invariant measurement and a representation for the lossless sample. ($(-\gamma)^{\frac{1}{2}}$ is positive real).

[14] The ambiguity in the sign of S_{11} arises from the $(\sigma_{11}+\sigma_{22})/2$ term in the exponential, where both σ_{11} and σ_{22} are indeterminate to within additive multiples of 2π. Inspection of Eq. (4) indicates that the two values of $\mathcal{Y}_\epsilon$ obtainable from the two values of S_{11} are reciprocals of each other. The choice of the proper sign of S_{11} can therefore be governed by the restriction placed on $\mathcal{Y}_\epsilon$ by Eq. (3), namely, that $Im\ (\mathcal{Y}_\epsilon)$ must be negative. It follows, of course, that $Im\ (1/\mathcal{Y}_\epsilon)$ can only be positive. One can therefore find $\mathcal{Y}_\epsilon$ from Eq. (4) without regard to sign and, if $Im\ (\mathcal{Y}_\epsilon)$ turns out to be positive, take $\mathcal{Y}_\epsilon$ as the reciprocal of the complex number found.

found first via the tangent relation method. Considerations similar to those above lead to the result

$$D_0' = [(D_0+S_0)+L-L_\epsilon]/2 \pm \lambda_g/8, \qquad (17)$$

where the ambiguity in sign is similar to the one discussed in reference 14. The dielectric constant is obtained from the admittance determinant

$$\mathcal{Y}_\epsilon = -\frac{1+\gamma \tan^2(2\pi D_0'/\lambda_g)}{\tan^2(2\pi D_0'/\lambda_g)+\gamma}, \qquad (18)$$

where that sign in Eq. (17) which makes $\mathcal{Y}_\epsilon > 1$ must be chosen. The fact that (D_0+S_0) appears as a sum is fortunate since this sum can often be found more accurately than its component parts. This measurement also may be simplified when D and S are both referred to the same plane T, as was done in connection with with Fig. 4, so that L in Eq. (17) will be identically zero.

ACKNOWLEDGMENT

The work described in this report was conducted under Contract AF-19(604)-890 sponsored by the Air Force Cambridge Research Center.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 2 FEBRUARY, 1955

Information Theory Aspects of Propagation through Time-Varying Media

JOSEPH FEINSTEIN*
National Bureau of Standards, Washington 25, D. C.
(Received February 27, 1954)

The channel capacity of a communications system which utilizes wave propagation through a time-varying medium such as the ionosphere or troposphere is evaluated in terms of the statistical properties of the medium and of the noise. The signal fading in such a system reduces the capacity. Rayleigh fading is found to give rise to an equivalent signal to noise ratio of 1.72, while shallow fading of the Gaussian type augments the noise in the channel by a fraction of the signal power proportional to the fading depth. An optimum manner of band width subdivision is shown to exist when selective fading is present.

Information theory concepts are broadened to include the possibility of multiple reception at spaced receiving sites, and the consequent increase in theoretical channel capacity is computed as a function of the number of such sites and the signal statistics. The method of maximum liklihood is utilized to obtain optimum combinatorial laws for the multiple signals. The commonly employed maximum signal selection diversity system is shown to perform as well as the optimum system in the presence of Rayleigh fading, for a small number of receiving sites.

INTRODUCTION

THE conventional treatment of the capacity of a communications channel given by information theory postulates steady signals contaminated by noise during the transmission process. In a wide class of systems, however, there is another effect which tends to degrade information: the signal fading associated with transmission through a time-varying medium. This is present, for example, in all systems utilizing space propagation of waves through the ionosphere or troposphere.

To compensate, as it were, for this effect the relatively large region of space occupied by the signal field when a propagation link is utilized in a communications system introduces additional degrees of freedom into the reception process. While semiempirical methods of taking advantage of these degrees of freedom, known as diversity, have long been known in the art, no standards are available against which their performance may be judged. It is the general purpose of this paper to present a fundamental approach to the problem of transmission through time-varying media, utilizing the formalism of information theory. The problems treated have been limited to calculations of channel capacity, rather than to a consideration of the effects of fading upon specific modulation systems. The results serve, however, to indicate the fundamental limitations present in this type of communications link.

In part I channel capacity is evaluated for the conventional single point reception of a fading signal in the presence of noise; several types of fading distributions and fluctuation characteristics are considered. The generalization to multiple receiving elements is developed in part II and applied to an examination of directional and diversity reception.

PART I. SINGLE POINT RECEPTION

1. Nature of the Received Signal

The physical distinction between fading and noise as mechanisms of signal degradation is evident from the fact that noise is received even in the absence of a transmitted signal while fading effects act to modulate the signal. The resultant amplitude received, y, when

* Now at Bell Telephone Laboratories, Murray Hill, New Jersey.

signal A is sent is of the form

$$y=AM+N, \tag{1}$$

where M is the fading modulation factor and N is the noise. The statistical characteristics of the random variable M are, of course, related to the fluctuations of the transmission medium. The conditions which the probability distribution $p(M)$ must satisfy are

$$\int_{-\infty}^{\infty} p(M)dM=1 \tag{2a}$$

and

$$\int_{-\infty}^{\infty} M^2p(M)dM=1. \tag{2b}$$

The first follows from the definition of probability and the second from the requirement of constant average power. If we take $p(M)$ to be of normal form

$$p_M(M)=\frac{1}{m\sqrt{2\pi}}\exp\left[-\frac{(M-d)^2}{2m^2}\right], \tag{3}$$

then to satisfy (2b)

$$d=(1-m^2)^{\frac{1}{2}}. \tag{4}$$

For $m\ll 1$, $d\simeq 1$, and we obtain a Gaussian distribution with a mean of unity; for $m=1$,

$$p_M(M)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{M^2}{2}\right). \tag{5}$$

The first of these agrees with experimental observations on ionospheric signals under conditions of shallow fading,[1] while the second occurs under conditions of Rayleigh fading. In the intermediate region the normal form no longer holds but is replaced by an expression involving Bessel functions.[1] We shall restrict our applications to the two limiting forms above since they serve to illustrate the types of behavior obtained and possess the merit of mathematical simplicity.

Consideration of the remaining property of M which we shall require, its rate of fluctuation with respect to the sequence of signals which constitute a message, will be deferred until section 5.

We make the conventional Gaussian white noise assumptions for N, so that

$$p_N(N)=\frac{1}{\sigma_N\sqrt{2\pi}}\exp\left(-\frac{N^2}{2\sigma_N{}^2}\right), \tag{6}$$

where $\sigma_N{}^2$ is the average noise power.

[1] R. W. E. McNicol, Proc. Inst. Elec. Engrs. (London) III **96**, 517 (1949).

2. Channel Capacity Formulation—Uncorrelated Fading

For a message consisting of n discrete signals we may write the expression of Shannon[2] for rate of information transmission in the form

$$C'=\int dA_1\int dA_2\cdots\int dA_n\cdot\int dy_1\int dy_2\cdots\int dy_n \cdot P(A_1,A_2\cdots A_n;y_1,y_2\cdots y_n) \cdot\log\frac{P(A_1,A_2\cdots A_n|y_1,y_2\cdots y_n)}{P(y_1,y_2\cdots y_n)}, \tag{7}$$

where conventional notation has been employed for the joint and conditional distributions. To convey maximum information the successive A's should be independent of each other. Since the noise contributions to each signal are also independent we may write the joint probability distribution of all the A's and y's as a product of individual distributions for each pair (A,y) provided the fading factor M is uncorrelated on successive signals. Under these conditions (7) reduces to

$$C'=n\int da\int dy\cdot P(A,y)\cdot\log[P(A|y)/P(y)]. \tag{8}$$

Making use of the relations

$$P(A,y)=p_A(A)\cdot P(A|y), \tag{9a}$$

$$P(y)=\int P(A,y)dA \tag{9b}$$

and

$$P(A|y)=\int p_M(M)\cdot p_N(y-AM)\cdot dM, \tag{10}$$

we may proceed to evaluate C' once we have selected specific forms for p_A and p_M. If the distribution function of the transmitted signal A is such as to maximize C', then $C'=C$, the channel capacity. For the case of a nonfading signal in noise it has been established[2] that a Gaussian p_A is optimum. It may be shown that this choice continues to yield maximum transmission rate in the case of shallow (displaced Gaussian) fading.† For Rayleigh fading the derivation of the maximal form of p_A, subject to the condition of constant variance is given in Appendix A; since this distribution is physically unrealizable we shall assume a Gaussian form for p_A here as well, so as to have a basis for intercomparison of our results. We may not, however, call C' the channel capacity in this case. We now proceed to evaluate C' for the two cases.

[2] C. E. Shannon, Bell System Tech. J. **27**, 633 (1948).

† The form of Eq. (12a) admits of this interpretation.

3. Evaluation of Capacity—Uncorrelated Fading

For Gaussian fading, (10) becomes

$$P(A|y)=\int_{-\infty}^{\infty}\frac{dM}{m\sqrt{2\pi}}\exp\left[-\frac{(M-1)^2}{2m^2}\right]\cdot\frac{1}{\sigma_N\sqrt{2\pi}}\exp\left[-\frac{(y-AM)^2}{2\sigma_N^2}\right] \quad (11)$$

$$=\frac{1}{[2\pi(\sigma_N^2+A^2m^2)]^{\frac{1}{2}}}\times\exp\left[-\frac{(y-A)^2}{2(\sigma_N^2+A^2m^2)}\right]. \quad (11a)$$

Substituting in (9), one obtains

$$P(y)=\int_{-\infty}^{\infty}\frac{dA}{\sigma_A\sqrt{2\pi}}\exp\left(-\frac{A^2}{2\sigma_A^2}\right)\cdot\frac{\exp\left[-\frac{(y-A)^2}{2(\sigma_N^2+A^2m^2)}\right]}{[2\pi(\sigma_N^2+A^2m^2)]^{\frac{1}{2}}}. \quad (12)$$

In view of the fact that $m\ll 1$ for this formulation to apply we assume $\sigma_N^2\gg m^2\sigma_A^2$, i.e., the effect of noise predominates over that of fading. Then an approximate evaluation of (12) leads to

$$P(y)=\frac{1}{\{2\pi(\sigma_N^2+\sigma_A^2[m^2+1])\}^{\frac{1}{2}}}\times\exp\left\{-\frac{y^2}{2(\sigma_N^2+\sigma_A^2[m^2+1])}\right\}^{\frac{1}{2}}. \quad (12a)$$

If we insert these quantities into (8), C' is given to the same order of approximation by

$$C'=\frac{n}{2}\log\left(1+\frac{\sigma_A^2}{\sigma_N^2+m^2\sigma_A^2}\right). \quad (13)$$

The form of this result enables a simple interpretation to be drawn of the effect of shallow fading, *viz.*, that the effective noise power is increased by $m^2\times$the average signal power, the total effective signal power remaining unchanged.

For Rayleigh fading, (10) becomes

$$P(A|y)=\int_{-\infty}^{\infty}\frac{dM}{\sqrt{2\pi}}\exp-\frac{M^2}{2}\cdot\frac{1}{\sigma_N\sqrt{2\pi}}\times\exp\left[-\frac{(y-AM)^2}{2\sigma_N^2}\right] \quad (14)$$

$$=\frac{1}{[2\pi(\sigma_N^2+A^2)]^{\frac{1}{2}}}\exp\left[-\frac{y^2}{2(\sigma_N^2+A^2)}\right]. \quad (14a)$$

By substituting in (9), there results

$$P(y)=\int_{-\infty}^{\infty}\frac{dA}{\sigma_A\sqrt{2\pi}}\exp\left(-\frac{A^2}{2\sigma_A^2}\right)\cdot\frac{\exp\left[-\frac{y^2}{2(\sigma_N^2+A^2)}\right]}{[2\pi(\sigma_N^2+A^2)]^{\frac{1}{2}}}. \quad (15)$$

If one were to assume the noise power to be much greater than the signal power, the same type approximation as was made in (12) could be introduced here, and would lead to the result $C'=0$. However, since such an approximation is valid only for signal to noise ratios much less than unity, it is of dubious interest. Instead we make the more physically interesting assumption that the noise power is much less than the signal power, so that we may set $\sigma_N^2=0$ and solve for the effect of fading alone in the absence of noise. Equation (15) then yields

$$P(y)=\frac{1}{\pi\sigma_A}K_0\left(\left|\frac{y}{\sigma_A}\right|\right), \quad (15a)$$

where K_0 is the modified Hankel function. Evaluation of (8) then yields a pure numeric,‡ independent of the signal power as one might expect: $C'/n=0.500$. One may express the effect of fading upon the rate of information transmission in terms of the signal-to-noise ratio which yields the same value of capacity by equating this result to $\frac{1}{2}\log[1+(\sigma_A^2/\sigma_N^2)]$, noting that natural base logarithms are to be employed. This procedure gives $(\sigma_A/\sigma_N)^2=1.72$ as the equivalent signal to noise power ratio.

The assumption of uncorrelated M values for neighboring signals made in this section, represents the worst possible situation from the point of view of degradation of information. We now proceed to consider the case where M possesses some degree of correlation from point to point and then to inquire into the factors which determine this correlation.

4. Evaluation of Capacity—Correlated Fading

In the general case we may write for the conditional probability distribution of the A's and y's

$$P(A_1,A_2\cdots A_n|y_1,y_2\cdots y_n)=\int dM_1\cdot p_M(M_1)\cdot p_N(y_1-A_1M_1)\cdot\int dM_2\cdot p_{M_2}(M_1|M_2)\cdot p_N(y_2-A_2M_2)\cdots\times\int dM_n\cdot p_{M_n}(M_1M_2\cdots M_{n-1}|M_n)\cdot p_N(y_n-A_nM_n), \quad (16)$$

‡ The $\int_0^\infty K_0(x)\log K_0(x)dx$ which arises here was evaluated numerically as -0.223.

where the p_{M_k} expresses the dependence of a given M_k upon the preceding $(k-1)M$'s. It seems reasonable to take a multivariate normal form for the conditional distribution of M. In view of the complexities attending such a choice we make the further assumption that M is a first-order Markoff process, thus restricting ourselves to bivariate distributions. For Gaussian fading, then,[3]

$$p_{M_k}(M_1,M_2,\cdots M_{k-1}|M_k)=p_M(M_{k-1}|M_k)$$
$$=\frac{\exp\left[-\dfrac{[M_k-1-\rho(M_{k-1}-1)]^2}{2m^2(1-\rho^2)}\right]}{m[2\pi(1-\rho^2)]^{\frac{1}{2}}}, \quad (17)$$

where ρ is the correlation coefficient between successive M's. The evaluation of (16) with the form (17) has been relegated to Appendix B. For $\rho\ll 1$, one obtains a reduction in fading variance m^2, by the factor $(1-\rho^2)$. Difficulties of integration have prevented evaluation for other ranges.

For $\rho\sim 1$, a possible choice for p_{M_k} may be obtained by taking M to be perfectly correlated in groups of q points, and uncorrelated between these groups. Utilizing the delta function the following representation is obtained for p_{M_k}:

$$P(A_1,A_2\cdots A_n|y_1,y_2\cdots y_n)$$
$$=P(A_1\cdots A_q|y_1\cdots y_q)$$
$$\cdot P(A_{q+1}\cdots A_{2q}|y_{q+1}\cdots y_{2q})\cdots, \quad (18)$$

$$P(A_1\cdots A_q|y_1\cdots y_q)$$
$$=\int dM_1\cdot p_N(M_1)\cdot p_N(y_1-A_1M_1)\cdots$$
$$\times\int dM_q\delta(M_q-M_1)\cdot p_N(y_q-A_qM_q). \quad (19)$$

For Gaussian fading an approximate evaluation of (19) leads for large q to a form similar to that obtained for uncorrelated fading but with an effective noise power given by (see Appendix C)

$$\sigma_N'^2=\sigma_N^2+m^2\sigma_A^2/1+m^2(q-1)(\sigma_A^2/\sigma_N^2). \quad (20)$$

The channel capacity is then found from (13) by relacing the fading depth factor m^2, by $m^2/1+m^2(q-1)(\sigma_A^2/\sigma_N^2)$. The value to be chosen for q is related in a general way to the rate at which the correlation curve falls off with signal element separation.

For the case of correlated Rayleigh fading, if the effect of the fading is reduced below that of the noise as a result of the correlation $[\sigma_N^2\gg\sigma_A^2(1-\rho^2)]$, then the above expressions are applicable, with $m=1$. If, however, the fading still remains the dominant factor in degrading information, we encounter some difficulty. For if noise is neglected then the δ-function formulation of the preceding paragraph leads to an infinite channel capacity. This is not surprising in view of the infinitely fine gradations which may be distinguished in the ratio of two completely unperturbed signals.

For a bivariate normal distribution of p_{M_k} (neglecting noise) the correlated Rayleigh fading leads to

$$P(A_1,A_2\cdots A_n|y_1,y_2\cdots y_n)$$
$$=\frac{1}{A_1\sqrt{2\pi}}\exp\left(-\frac{y_1^2}{2A_1^2}\right)$$
$$\cdot\frac{\exp\left\{-\dfrac{[(y_2/A_2)-\rho(y_1/A_1)]^2}{2(1-\rho^2)}\right\}}{A_2[2\pi(1-\rho^2)]^{\frac{1}{2}}}\cdots$$
$$\times\frac{\exp\left\{-\dfrac{[(y_n/A_n)-\rho(y_{n-1}/A_{n-1})]^2}{2(1-\rho^2)}\right\}}{A_n[2\pi(1-\rho^2)]^{\frac{1}{2}}}. \quad (21)$$

Difficulties of integration prevent us from obtaining $P(y_1,y_2\cdots y_n)$ explicitly. For $\rho\ll 1$, however, it is reasonable to assume that this y-distribution is not materially affected by the correlation. The increase in transmission rate over the uncorrelated case then takes the form

$$\Delta C'=\left(\frac{n-1}{2}\right)\log(1/1-\rho^2). \quad (22)$$

5. Nature of Fading Correlation

In the preceding section we have introduced the symbol ρ to denote the correlation of the fading variable M between adjacent signal elements of a message. Now it is well known[2] that a message of duration T, and occupying band width W, requires $2TW$ values, or signal elements for its specification. Consequently the signals corresponding to such a message are spaced at time intervals $1/(2W)$. Then if M does not vary with frequency over the band W, ρ is simply the autocorrelation function of M, evaluated for a time displacement $1/(2W)$. In the more general case, however, M may vary with frequency as well as with time; we therefore define

$$C_0(\omega,\tau)=\lim_{T\to\infty}\frac{1}{T}\int_0^T M(\omega_0,t)\cdot M(\omega_0+\omega,\ t+\tau)\cdot dt. \quad (23)$$

One may obtain a crude value of ρ when selective fading is present by dividing the band W into segments of width p_0, where p_0 is the frequency interval over which the fading remains fairly well correlated. Then in each of the subdivisions the sampling rate is $2p_0$ signals/second, so that ρ is again the autocorrelation function of M, but now evaluated for a time displacement $1/(2p_0)$. Since this time interval is greater than $1/(2W)$, ρ will

[3] J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals* (McGraw-Hill Book Company, Inc., New York, 1950), p. 49.

decrease as expected. A more accurate approach has as its point of departure, the Fourier expression for a signal as the sum of its frequency components:

$$S(t)=\sum_k a_k \cos(\omega_k t+\phi_k), \tag{24}$$

where the frequencies are spaced at intervals $1/T$. For subdivisions of band width p, ρ is given by

$$\rho_p=\lim_{T\to\infty}\frac{1}{T}\int_0^T dt\cdot\frac{S(t)\cdot M(\omega,t)\cdot S(t')\cdot M(\omega,t')}{S(t)\cdot S(t')}, \tag{25}$$

where $t'=t+(1/2p)$. When (24) is inserted, there results

$$\rho_p=\frac{1}{T}\int_0^T dt \cdot\frac{\sum_k\sum_l\{M(\omega_k,t)\cdot M(\omega_l,t')[a_k\cos(\omega_k t+\phi_k)]}{\sum_k\sum_l\{[a_k\cos(\omega_k t+\phi_k)]}\cdot\frac{[a_l\cos(\omega_l t'+\phi_l)]\}}{[a_l\cos(\omega_l t'+\phi_l)]\}}. \tag{25a}$$

Since the ϕ_k are random and independent for transmission of maximum information, the sum of cross-product terms arising from $\sum_k\sum_{l\neq k}$ may be taken as zero. Then

$$\rho_p=\frac{\sum_k C_0(\omega_k,\ \tau=1/2p)\cdot\cos(\omega_k/2p)}{\sum_k\cos(\omega_k/2p)}. \tag{26}$$

Replacing the summation by an integration as we pass to the limit $T=\infty$ and maximizing with respect to p, we obtain

$$\rho=\operatorname*{Max}_{(p)}\frac{\int_{-p/2}^{p/2}C_0(\omega,\ \tau=1/2p)\cdot\cos(\omega/2p)\cdot d\omega}{\int_{-p/2}^{p/2}\cos(\omega/2p)\cdot d\omega}. \tag{26a}$$

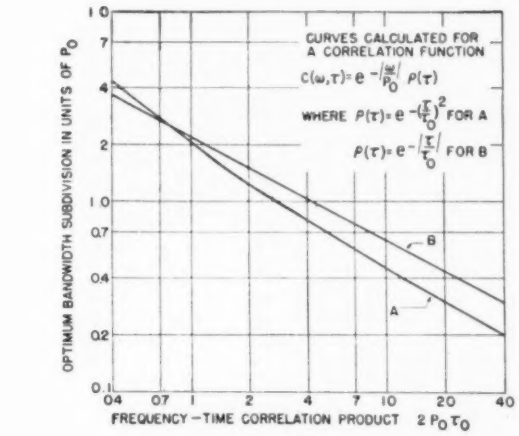
As an example we take

$$C_0(\omega,\tau)=\exp(-\tau^2/\tau_0^2)\cdot\exp(-|\omega/p_0|) \tag{27}$$

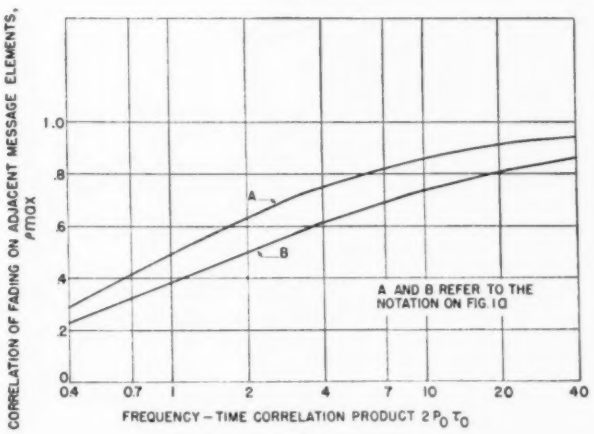
and let $p=xp_0$; then

$$\rho(x)=\exp\left[-\left(\frac{1}{2p_0\tau_0 x}\right)^2\right]\cdot\frac{1}{1+4x^2}\times\left[\frac{2x}{\sin\frac{1}{4}}+e^{-x/2}-2xe^{-x/2}\cot\tfrac{1}{4}\right]. \tag{28}$$

The optimum fraction x is plotted as a function of the $2p_0\tau_0$ product in Fig. 1a, and the corresponding maximum ρ is shown in Fig. 1b. Similar curves are given for an M correlation of the form $C_0(\omega,\tau)=e^{-|\tau/\tau_0|}\cdot e^{-|\omega/p_0|}$ to



(a)


CORRELATION OF FADING ON ADJACENT MESSAGE ELEMENTS, ρ_{TOX}
A AND B REFER TO THE NOTATION ON FIG. 1C
FREQUENCY – TIME CORRELATION PRODUCT $2P_0\tau_0$
A
B

(b)

FIG. 1. (a). Optimum band width subdivision in the presence of selective fading. (b). Fading correlation for the optimum case.

illustrate the dependence of the results on the specific form assumed. These curves make it clear that there is an optimum manner of subdivision of a band $W>p_0$ for a given set of fading characteristics. It is also to be noted that as $2p_0\tau_0$ falls below unity, $\rho\to 0$, even for optimum subdivision. Consequently the results of section 1, which assumed no correlation between adjacent points, are valid for $2p_0\tau_0$ somewhat less than one.

Such small values of $p_0\tau_0$ raise certain questions regarding the operational significance of the definition (23), wherein an envelope comparison of two continuous signals, initially transmitted on frequencies ω_0 and $\omega_0+\omega$, is visualized. This requires that the band width of the receivers be sufficient to admit the modulation frequencies produced by the fluctuation of M. But for a temporal correlation interval τ_0, modulation frequencies of order $1/\tau_0$ are present. Consequently separation of the two signals becomes uncertain when their frequency difference ω falls to $1/\tau_0$; thus it becomes impossible to distinguish frequency correlation intervals $p_0<1/\tau_0$ if both signals are received at the same point in space. Conceptually, however, it is possible to visualize an ensemble of identically fluctuating systems so that the

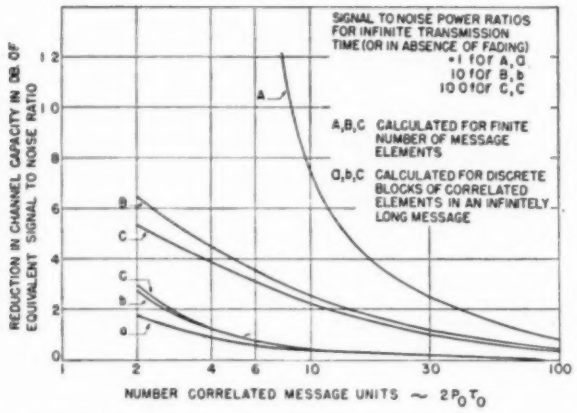

SIGNAL TO NOISE POWER RATIOS FOR INFINITE TRANSMISSION TIME (OR IN ABSENCE OF FADING)
=1 for A, a
10 for B, b
100 for C, c
A,B,C CALCULATED FOR FINITE NUMBER OF MESSAGE ELEMENTS
a,b,c CALCULATED FOR DISCRETE BLOCKS OF CORRELATED ELEMENTS IN AN INFINITELY LONG MESSAGE
REDUCTION IN CHANNEL CAPACITY IN DB OF EQUIVALENT SIGNAL TO NOISE RATIO
NUMBER CORRELATED MESSAGE UNITS ~ $2P_0T_0$

FIG. 2. Reduction in channel capacity produced by a finite number of message elements.

signals are separated physically. No limit would then exist on the detectable frequency separation. In any event the exact values of p_0 and τ_0 are no longer of importance once their product has fallen below unity. It is interesting to note that $2p_0\tau_0 \sim 1$ has been identified by Gabor[4] with a quantum, or elementary cell of information.

6. Finite Transmission Time Viewpoint

To realize channel capacity, infinite transmission time is required. It seems reasonable to view fading as destroying the channel coherence periodically, and therefore as acting to shorten the effective time of transmission of a message. One would expect this viewpoint to be applicable only when fading is not the dominant factor limiting channel capacity and to be related to the delta function correlation assumption in Sec. 4.

The following treatment takes as its point of departure the expression given by Shannon[5] for the probability of uncertainty of each bit of a received message when a finite transmission time is employed:

$$p=2^{-T\eta}, \tag{29}$$

where $\eta=C-R$, C being the channel capacity and R the actual rate of information transmission. The equivocation produced by this uncertainty,

$$H=-\sum_n p_n \log p_n, \tag{30}$$

where the summation is over all the possible states of the system, represents the diminution in transmission rate and is therefore equal to η, when expressed on a per second basis:

$$\eta=-2W[p\log p+(1-p)\log(1-p)]. \tag{31}$$

Inserting (29), and setting $\eta=2Wg$, so that g is reckoned on a bit basis, we find

$$g=[(2TW)g\cdot 2^{-(2TW)g} - (1-2^{-(2TW)g})\log_2(1-2^{-(2TW)g})]. \tag{32}$$

We may now insert the correlation interval $\tau_0 p_0$ for TW, and solve this transcendental equation numerically for g as a function of $2\tau_0 p_0$. The results have been plotted in Fig. 2 in terms of the db reduction in signal-to-noise ratio below that of an ideal channel. Equation (20), with $m=1$ and $q=2\tau_0 p_0$, has been similarly plotted for comparison. The nature of the reasoning leading to (29) is such that one would expect the results to be meaningful only for fairly large $\tau_0 p_0$. Comparison of these curves indicates that the two effects are only qualitatively similar. The dissimilar dependence upon initial signal to noise ratio is especially noteworthy.

PART II. MULTIPLE ELEMENT RECEPTION

1. Space Characteristics of the Signal Field

We turn now to systems of information transmission which utilize relatively unguided wave propagation through unconfined media. To evaluate the performance of such systems requires a generalization of the treatment in part I. To this end it is desirable to introduce the notion of angular power spectrum of received energy, and the related concept of a ground correlation pattern.[6] Briefly, the medium inhomogeneities whose fluctuations give rise to fading also tend to destroy the phase coherence which is responsible for confining the received energy to a cone of angle determined by the Fresnel zone geometry for the system. As a consequence, power may be received over a very much wider and fluctuating angle when such inhomogeneities are present; the angular power spectrum $P(\theta)$ gives the time average power received at angle θ, measured from the specular direction.

Because of the finite distance over which the fluctuations within the medium are correlated, the correlation between signals received at any two points near the ground will fall as the points are further separated. The correlation of the signals as a function of their spacing, known as the ground correlation pattern, has been shown to be[6] the Fourier transform of the angular power spectrum. For an antenna with pattern factor $F(\theta)$ set noise N_s and antenna noise field N_a/unit solid angle (assumed isotropic) the signal to noise ratio is given by

$$S/N=\frac{\int P(\theta)\cdot F(\theta)\cdot d\Omega}{N_s+N_a\int F(\theta)d\Omega}. \tag{33}$$

For antenna noise predominant over set noise this has a maximum value P_m/N_a, where P_m is the maximum power density in the spectrum. In practice of course there is not much point in obtaining greater directionality than that which corresponds to the angle between, say, the half power points of $P(\theta)$. The reason little further gain is obtainable is directly related to the fact

[4] D. Gabor, J. Inst. Elec. Engrs. (London) III **93**, 429 (1946).
[5] See reference 2, p. 413.
[6] J. Feinstein, Trans. Inst. Radio Engrs. PGAP **AP-2**, 23 (1954).

that a more extensive structure will occupy a region greater than that over which the signal remains correlated, so that phase incoherence sets in. A new effect then arises, termed diversity reception.

We visualize the simultaneous reception of the transmitted signal on a set of elements, each of which may be a directional antenna (for optimum performance), the spacing between elements being greater than the correlation distance at the receiving site. Consequently the fading of the signals received on the individual elements is uncorrelated. Since the correlation distance of the noise field is generally much smaller than that of the signal field, the noise component of the signals is also assumed uncorrelated. In this connection it should be noted that the use of a directional antenna may alter the fading characteristics of a received signal. A well-known example of such behavior is found in the so-called MUSA antenna, employed to select a single ionospheric mode so as to reduce the selective fading which arises from mode interference. This situation corresponds to the existence of several maxima in $P(\theta)$, one of these being selected by $F(\theta)$.

In the next section we deduce the increase in channel capacity which results from multiple reception, with our two types of fading, and with various resultant signal selection laws.

2. Channel Capacity with Diversity Reception

Given a set of l independent signals: y_1, $y_2 \cdots y_l$ corresponding to a transmitted signal A, we are faced with the problem of selecting a method of combining the individual signals. It would be desirable to choose this law of combination so as to minimize our error of determination of A. This problem is well known in statistics; for minimization of the variance of the distribution of the combined signal $\bar{A}$, the method of maximum likelihood is appropriate.[7] This method yields for a Gaussian distribution (displaced from the true value A) of the individual signals

$$\bar{A}=\frac{1}{l}\sum_{k}^{l} y_k \tag{34}$$

i.e., the arithmetic mean of the instantaneous signals. For the case of Rayleigh fading accompanied by noise, as specified by Eq. (14a), one finds

$$\bar{A}^2=\frac{1}{l}\sum_{k}^{l} y_k^2-\sigma_N^2 \tag{35}$$

a root mean square or demodulated type of combination. We shall examine diversity systems obeying these optimum combinational laws, as well as certain others which are of interest.

We consider first the case in which each received message element is taken as the arithmetic mean $\bar{y}_l$, of the l individual signals y_1, $y_2 \cdots y_l$ corresponding to a single transmitted signal A. Then for a single message element

$$P(A\,|\,\bar{y}_l)=\int dy_1 \int dy_2 \cdots \int dy_{l-1}\cdot p(A\,|\,y_1) \cdot p(A\,|\,y_2)\cdots p(A\,|\,l\bar{y}_l-[y_1+y_2+\cdots y_{l-1}]). \tag{36}$$

For Gaussian fading this yields

$$P(A\,|\,\bar{y}_l)=\frac{1}{\left[2\pi\dfrac{(\sigma_N^2+A^2m^2)}{l}\right]^{\frac{1}{2}}} \times\exp\left[-\frac{l(\bar{y}_l-A)^2}{2(\sigma_N^2+A^2m^2)}\right], \tag{37}$$

where (11a) has been employed for each $p(A/y)$. The reduction in variance by $\sqrt{l}$ leads to the increased capacity

$$C_l'=\frac{n}{2}\log\left(1+\frac{\sigma_A^2 l}{\sigma_N^2+m^2\sigma_A^2}\right) \tag{38}$$

for uncorrelated fading on successive transmitted message elements. This result may be viewed as a decrease in effective noise power by the factor l.

Applying (36) to Rayleigh fading we get a similar reduction in variance

$$P(A\,|\,\langle y_l^2\rangle)=\frac{\sqrt{l}}{A\sqrt{2\pi}}\exp\left(-\frac{l\langle y_l^2\rangle}{2A^2}\right), \tag{39}$$

but this does not lead to any increase in channel capacity because the transformation $u=y\sqrt{l}$ leaves the expression for C invariant. Physically this result is traceable to the mean value of zero associated with this type of distribution.

Using the optimum combinatorial law for Rayleigh fading specified by (35), with $\sigma_N^2=0$, we obtain for the arithmetic mean of square law demodulated signals[8]

$$P(\bar{X}_l)=\frac{(l/2)^{l/2}}{\Gamma(l/2)}\exp\left(-\frac{l\bar{X}_l}{2}\right)\cdot\bar{X}_l^{(l/2)-1}, \quad \bar{X}_l\geq 0. \tag{40}$$

The conditional distribution $P(A\,|\,\langle y_l\rangle^2)$ is obtained from (40) by setting $\bar{x}_l=\langle y_l^2\rangle/A^2$. The distribution of $\langle y_l^2\rangle$ becomes

$$P\langle y_l^2\rangle=\int_{-\infty}^{\infty}\frac{dA}{\sigma_A\sqrt{2\pi}}\exp\left(-\frac{A^2}{2\sigma_A^2}\right)\cdot\frac{(l/2)^{l/2}}{A^2\Gamma(l/2)} \times\exp\left(-\frac{l\langle y_l^2\rangle}{2A^2}\right)\cdot\left(\frac{\langle y_l^2\rangle}{A^2}\right)^{(l/2)-1} \tag{41}$$

$$=\frac{l^{(l/4)}}{2^{(l/2)}\Gamma(l/2)}\exp\left[-\frac{1}{\sigma_A}(l\langle y_l^2\rangle)^{\frac{1}{2}}\right] \times\frac{\langle y_l^2\rangle^{(l/4)-1}}{(\sigma_A)^{l/2}}\cdot F_l\frac{\langle y_l^2\rangle}{A}, \tag{41a}$$

[7] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey, 1946), p. 498.

[8] See reference 7, p. 236.

TABLE I.

l	Cunits/message element
1	0.500
2	0.693
4	1.01
6	1.20

for l even, where $F_2(x)=1$, $F_4(x)=1+(1/x)$, $F_6(x)=(3/x)+(3/x^2)+1$. The integrals which arise in determining C' can be evaluated[9] in terms of the gamma function and its logarithmic derivative, plus some numerical integration occasioned by the presence of the F function. It is because of the labor involved in these numerical integrations that l has been taken up to six only. The resultant values of transmission rate are tabulated below; for comparison, the single element case which as might be expected yields the same value of C' for the variable y_1^2 as for y_1, the variable considered in part I, has been listed in Table I. For large values of l we may make use of the central limit theorem to replace (41) by the Gaussian distribution

$$P(\bar{X}_l)\simeq\left(\frac{l}{4\pi}\right)^{\frac{1}{2}}\exp\left[-\frac{l}{4}(\bar{X}_l-1)^2\right], \tag{42}$$

so that

$$P\langle y_l^2\rangle=\int_{-\infty}^{\infty}\frac{dA}{\sigma_A\sqrt{2\pi}}\exp\left(-\frac{A^2}{2\sigma_A{}^2}\right)\frac{1}{A^2}\left(\frac{l}{4\pi}\right)^{\frac{1}{2}}\times\exp\left[-\frac{l}{4}\frac{(\langle y_l^2\rangle-A^2)^2}{A^4}\right]; \tag{43}$$

for $l\gg1$,

$$P\langle y_l^2\rangle\simeq\frac{\langle y_l^2\rangle^{-\frac{1}{2}}}{\sigma_A\sqrt{2\pi}}\exp\left(-\frac{\langle y_l^2\rangle}{2\sigma_A{}^2}\right), \tag{43a}$$

i.e., the distribution of $\langle y_l^2\rangle$ approaches that of A^2. The channel transmission rate under these conditions is given by

$$C'=\int_0^{\infty}d\langle y_l^2\rangle\int_{-\infty}^{\infty}\frac{dA}{\sigma_A\sqrt{2\pi}}\exp\left(-\frac{A^2}{2\sigma_A{}^2}\right)\cdot\frac{1}{A^2}\left(\frac{l}{4\pi}\right)^{\frac{1}{2}}\exp\left[-\frac{l}{4}\left(\frac{\langle y_{l2}\rangle}{A^2}-1\right)^2\right]\cdot\log\frac{\left(\frac{l}{4\pi}\right)^{\frac{1}{2}}\frac{1}{A^2}\exp\left[-\frac{l}{4}\left(\frac{\langle y_l^2\rangle}{A^2}-1\right)^2\right]}{\frac{1}{\sigma_A(2\pi\langle y_l^2\rangle)^{\frac{1}{2}}}\exp\left(-\frac{\langle y_l^2\rangle}{2\sigma_A{}^2}\right)} \tag{44}$$

$$=\tfrac{1}{2}\log l+\frac{\epsilon}{2}, \tag{44a}$$

where ϵ is Euler's constant. The equivalent signal to noise ratio increases approximately linearly with l, the same type of increase found for the instantaneous mean of Gaussian fading. The corresponding increase in db of effective signal to noise ratio has been plotted as a function of l in Fig. 3 for the two types of fading.

Finally, we consider the diversity system commonly employed today; this system selects for each message element the maximum signal present in the set of receiving elements. The conditional probability distribution of this signal, denoted by $\ddot{y}_l$, is for Rayleigh fading

$$P(A\,|\,\ddot{y}_l)=\frac{l}{A\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\left(\frac{\ddot{y}_l}{A}\right)^2\right]\times\left[\operatorname{erf}\left(\frac{\ddot{y}_l}{A\sqrt{2}}\right)\right]^{l-1}. \tag{45}$$

It is possible to expand the error function in a Maclaurin series of which only the first term need be retained if l is not too large. Consequently we replace (45) with

$$P(A\,|\,\ddot{y}_l)\simeq\left(\frac{\ddot{y}_l}{A\sqrt{2}}\right)^{l-1}\frac{\exp\left[-\frac{1}{2}\left(\frac{\ddot{y}_l}{A}\right)^2\right]}{A\cdot\sqrt{2}\cdot\Gamma(l/2)}. \tag{45a}$$

With the substitution $u=l\ddot{y}_l$, (45a) becomes identical with (41). Consequently the transmission rates of the two systems are equal for corresponding values of l, so long as the approximation of (45) by (45a) holds. This appears to be valid through $l=4$. For higher values of l a decrease in performance as compared to the optimum system specified by (35) is to be expected, and for very large values of l the asymptotic efficiency in the sense defined by Cramér approaches zero.

A word should be said regarding the relative effects of noise upon these two types of diversity. The maximum signal selection system is preferred in practice because of the deleterious effects of noise upon a straight average of

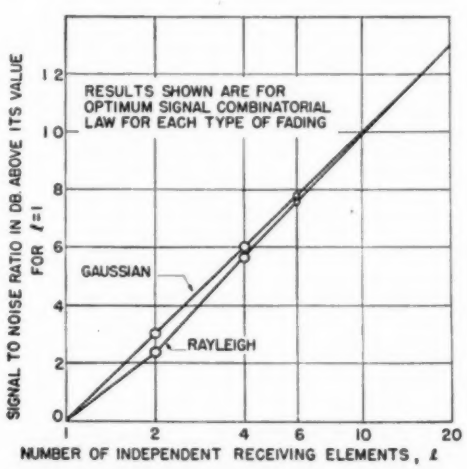


FIG. 3. Increase in effective signal-to-noise ratio produced by diversity reception.

[9] W. Grobner and N. Hofreiter, *Integraltafel* (Springer Publishing Company, Inc., Vienna, 1950), part II.

demodulated signals. To the author's knowledge there is no system in use which attempts to subtract off this noise in accordance with the prescription given by (35), although it may be that the variability with time of σ_N^2 makes this course impractical. In any event it is apparent that the performance of the maximum selection system will deteriorate rapidly once signal to noise ratios in the vicinity of unity are reached.

CONCLUSION

In this paper it is hoped a start has been made on the analysis of communications systems in which signal fading occurs. The two types of fading distributions used, Rayleigh and Gaussian, were chosen to illustrate situations wherein fading is, respectively, a dominant and a minor factor; subsequent approximations made in carrying through the analysis were consistent with this aim. The fundamental viewpoint of information theory was adopted throughout, so that consideration has been limited to channel transmission rate calculations.

For shallow fading it has been shown that the signal fluctuations give rise to an equivalent noise power. A quantitative evaluation of the rather strong degradation of information produced by single point reception of a Rayleigh fading signal has been presented, and the amelioration possible with diversity systems has been explored. In this connection it has been shown that for a small number of receiving elements, the maximum signal selection type of diversity performs as well as the optimum, mean square type. The relation between the spectral and temporal characteristics of fading, and the correlation of the fading modulation on adjacent message elements has been developed, and has been shown to lead to an optimum manner of band subdivision. Expressions for the increase in transmission rate produced by partial correlation of the fading on successive signals have been developed for a variety of special cases, and from several viewpoints, although the rather intuitive justification for some of the approximations made make these results trustworthy only as order of magnitude estimates.

With the exception of section 6 of part I all values of transmission rate are based upon the natural logarithm as a measure of information. The more physically meaningful equivalent db of signal-to-noise ratio has been employed on the curves.

APPENDIX A

Derivation of p_A distribution which maximizes transmission rate in the presence of Rayleigh fading

$$C' = \int\int dA \cdot dy \cdot p_A(A) \cdot P(A|y) \log \frac{P(A|y)}{P(y)} \tag{1A}$$

$$= \int dA \cdot p_A(A) \cdot \int dy P(A|y) \log P(A|y) - \int dy \cdot P(y) \log P(y). \tag{2A}$$

In accordance with the methods of the calculus of variations, we add as constraints $-\mu \int p_A(A)dA - \lambda \int A^2 p_A(A)dA$, since we wish to maximize subject to the conditions

$$\int p_A(A)dA = 1 \quad \text{and} \quad \int A^2 p_A(A)dA = \sigma_A^2. \tag{3A}$$

For Rayleigh fading (in the absence of noise)

$$P(A|y) = \frac{1}{A\sqrt{2\pi}} \exp\left(-\frac{y^2}{2A^2}\right). \tag{4A}$$

Differentiating with respect to p_A, we find

$$\frac{\partial[C+\text{constraints}]}{\partial p_A} = \int dA(-\log A - \lambda A^2 - \mu) - \int dy[1+\log P(y)]\frac{\partial P(y)}{\partial p_A}, \tag{5A}$$

where

$$P(y) = \int \frac{dA}{A\sqrt{2\pi}} \exp\left(-\frac{y^2}{2A^2}\right) \cdot p_A(A), \tag{6A}$$

so that

$$\frac{\partial P(y)}{\partial p_A} = \int \frac{dA}{A\sqrt{2\pi}} \exp -\frac{y^2}{2A^2}. \tag{7A}$$

We now set (5A) equal to zero and interchange the order of integration in the last term on the right, ignoring questions of convergence for the moment.

$$\int dA\left[-\log A - \lambda A^2 - \mu - 1 - \frac{1}{A\sqrt{2\pi}} \times \int dy \exp\left(-\frac{y^2}{2A^2}\right) \log P(y)\right] = 0. \tag{8A}$$

This has the solution

$$\log P(y) = -\mu - 1 - \lambda y^2 - \log y - (1/2)(\epsilon + \log 2), \tag{9A}$$

so that

$$P(y) = \frac{c}{y} \exp(-\lambda y^2), \tag{10A}$$

where c and λ are to be chosen to satisfy Eq. (3A). The y distribution diverges when integrated over the infinite domain as a result of the $(1/y)$ factor. This is related to the interchange of limits performed above. We may proceed nevertheless by writing an exponent ν on this factor, and allowing ν to approach unity only in our final normalized expressions. To find p_A we must solve

$$\int_{-\infty}^{\infty} \frac{dA}{A\sqrt{2\pi}} p_A(A) \exp\left(-\frac{y^2}{2A^2}\right) = \frac{c}{y^\nu} \exp(-\lambda y^2). \tag{11A}$$

After the transformation $x=1/A^2$, the left-hand side takes the form of a Laplace transform

$$\frac{1}{\sqrt{2\pi}}\int_0^\infty dx\left[\frac{1}{x}p_A\left(\frac{1}{x}\right)\right]\exp\left[-\left(\frac{y^2}{2}\right)x\right]$$

$$=\frac{c}{y^\nu}\exp(-\lambda y^2). \quad (12A)$$

This has the solution

$$p_A(A)=\frac{c\sqrt{2\pi}}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)}\cdot\frac{1}{A^\nu(1-2\lambda A^2)^{1-\nu/2}},\quad |A|<\frac{1}{\sqrt{2\lambda}} \quad (13A)$$

$$=0,\qquad |A|>\frac{1}{\sqrt{2\lambda}}.$$

The conditions (3A) lead to the following values of the parameters in the vicinity of $\nu=1$:

$$\lambda=\frac{1}{\sigma_A{}^2\Gamma(1-\nu/2)};$$

$$c=\frac{1}{2\Gamma(1-\nu/2)}\cdot\left[\frac{2}{\sigma_A{}^2\Gamma(1-\nu/2)}\right]^{(1-\nu)/2}. \quad (14A)$$

The leading term in the expression for channel capacity is $\frac{1}{2}\log\Gamma(1-\nu/2)$ which becomes infinite for $\nu=1$. Since $\lambda\to 0$ under these conditions, the value of A having infinite probability, $A=1/\sqrt{2\lambda}\to\infty$. Consequently this distribution function is deemed physically unrealizable. It results from the lack of a constraint on the maximum value permitted for A.

APPENDIX B

Upon inserting Eq. (17) into Eq. (16) there results

$$P(A_1,A_2\cdots A_n|y_1,y_2\cdots y_n)$$
$$=\int_{-\infty}^{\infty}\frac{dM_1}{m\sqrt{2\pi}}\exp\left[-\frac{(M_1-1)^2}{2m^2}\right]\cdot\frac{1}{\sigma_N\sqrt{2\pi}}$$
$$\times\exp\left[-\frac{(y_1-A_1M_1)^2}{2\sigma_N{}^2}\right]\cdot\int_{-\infty}^{\infty}\frac{dM_2}{m[2\pi(1-\rho^2)]^{\frac{1}{2}}}$$
$$\times\exp\left\{-\frac{[M_2-1-\rho(M_1-1)]^2}{2m^2(1-\rho^2)}\right\}\cdots$$
$$\times\int_{-\infty}^{\infty}\frac{dM_n}{m\sqrt{2\pi(1-\rho^2)}}$$
$$\times\exp\left\{-\frac{[M_n-1-\rho(M_{n-1}-1)]^2}{2m^2(1-\rho^2)}\right\}. \quad (1B)$$

Integrating over M_n first we obtain the factor

$$\frac{\exp\left[-\frac{(y_n-A_n)^2}{2\sigma_N{}^2}\right]}{\{2\pi[\sigma_N{}^2+A^2m^2(1-\rho^2)]\}^{\frac{1}{2}}}$$
$$\times\exp\left\{\frac{\left[\rho\frac{(M_{n-1}-1)}{m(2(1-\rho^2))^{\frac{1}{2}}}+\frac{m(1-\rho^2)^{\frac{1}{2}}}{\sigma_N{}^2\sqrt{2}}A_n(y_n-A_n)\right]^2}{1+\frac{A_n{}^2m^2}{\sigma_N{}^2}(1-\rho^2)}\right\}. \quad (2B)$$

For $\rho\ll 1$, if we neglect the factor $\rho(M_{n-1}-1)$, then we obtain a simple Gaussian distribution for (y_n-A_n) with variance

$$\sigma^2=\sigma_N{}^2+m^2\sigma_A{}^2(1-\rho^2), \quad (3B)$$

and succeeding integrations are independent.

APPENDIX C

For Gaussian fading Eq. (19) takes the form

$$P=\int_{-\infty}^{\infty}\frac{dM}{m\sqrt{2\pi}}\exp\left[-\frac{(M-1)^2}{2m^2}\right]$$
$$\cdot\left(\frac{1}{\sigma_N\sqrt{2\pi}}\right)^q\exp\left[-\sum_{k=1}^{q}\frac{(y_k-MA_k)^2}{2\sigma_N{}^2}\right] \quad (1C)$$
$$=\left(\frac{1}{\sigma_N\sqrt{2\pi}}\right)^q\frac{1}{\left(1+\frac{m^2}{\sigma_N{}^2}\sum A_k{}^2\right)^{\frac{1}{2}}}$$
$$\times\exp\left\{-\frac{1}{2\sigma_N{}^2}\sum_k^q\left[(y_k-A_k)^2-\frac{\frac{m^2}{\sigma_N{}^2}A_k(A_k\cdots y_k)}{1+\frac{m^2}{\sigma_N{}^2}\sum_k A_k{}^2}\sum_q^l A_l(A_l-y_l)\right]\right\}. \quad (2C)$$

The double sum $\sum_k\sum_{l\neq k}A_kA_l(A_k-y_k)(A_l-y_l)$ tends to cancel if q is sufficiently large, since each of its individual factors is independent and random. There

remains

$$P=\left(\frac{1}{\sigma_N\sqrt{2\pi}}\right)^q \frac{1}{\left(1+\dfrac{m^2}{\sigma_N{}^2}\sum A_k{}^2\right)^{\frac{1}{2}}}$$

$$\times\exp\left\{-\frac{1}{2\sigma_N{}^2}\sum_k^q (y_k-A_k)^2\left[\frac{1+\dfrac{m^2}{\sigma_N{}^2}\sum\limits_{k'\neq k}^{q} A_{k'}{}^2}{1+\dfrac{m^2}{\sigma_N{}^2}\sum\limits_{k'} A_{k'}{}^2}\right]\right\}. \quad (3C)$$

For large q we may set

$$\sum_k^q A_k{}^2\simeq q\sigma_A{}^2;$$

this yields

$$P=\left(\frac{1}{\sigma_N(2\pi)^{\frac{1}{2}}}\right)^q \frac{1}{\left(1+m^2q\dfrac{\sigma_A{}^2}{\sigma_N{}^2}\right)^{\frac{1}{2}}}$$

$$\times\exp\left\{\frac{-\sum\limits_k^q (y_k-A_k)^2}{2\sigma_N{}^2\left[\dfrac{1+qm^2\sigma_A{}^2/\sigma_N{}^2}{1+(q-1)m^2\sigma_A{}^2/\sigma_N{}^2}\right]}\right\}. \quad (4C)$$

Since the normalizing factor $(1+m^2q\sigma_A{}^2/\sigma_N{}^2)^{\frac{1}{2}q}$ approaches unity at large q as does the bracketed exponential factor which modifies $\sigma_N{}^2$, the interpretation given by Eq. (20) of the text follows.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 2 FEBRUARY, 1955

Analog Computer Construction of Conformal Maps in Fluid Dynamics

N. P. Tomlinson, M. Horowitz, and C. H. Reynolds*
Aerophysics Departments, Goodyear Aircraft Corporation, Akron, Ohio
(Received April 12, 1954)

A general-purpose electronic analog computer has been used extensively to construct the streamlines around two-dimensional airfoil shapes immersed in an incompressible, nonrotating, nonviscous fluid. The computer generates a circular Lissajous pattern on a plotting table, and then transforms the coordinates of the circle into the airfoil shape by a conformal transformation. Two samples of the variety of airfoil shapes so constructed are shown. The accuracy of the process, which has been checked by a digital computer and a desk calculator, is quite satisfactory; the speed of the actual mapping process is highly advantageous to thorough exploration of a variety of airfoil shapes.

INTRODUCTION

CONFORMAL maps constructed with a general-purpose electronic analog computer of the Geda† type have been applied to two problems in fluid dynamics: (1) the determination of airfoil shapes; and (2) flow patterns of an incompressible, frictionless fluid. Meneley and Morrill[1] have given a description of this type of computer and its general uses.

Classical solutions of fluid-dynamic problems are based on Laplace's differential equation with specified boundary conditions. The solution of this equation with boundary conditions can be found by the mathematical technique of conformal mapping.

The principal use of conformal mapping is the transformation of a boundary-value problem in one domain to a corresponding problem in a simpler domain. For example, assume that Laplace's equation is to be solved, subject to certain boundary conditions within a region R. Unless R is a very simple shape, the solution by a direct method of attack will usually be difficult. However, by the use of a suitable transformation, it may be possible to convert R into some simpler region R_1, such as a circle or half-plane. If Laplace's equation can be solved in R_1 subject to the transformed-boundary conditions, then this solution (carried back to R by the inverse transformation) will be the required solution of the original problem. Transformation from one domain to another is easily completed with analog-computing equipment.

APPARATUS AND PROCEDURE

Conformal Mapping

The development of equations to determine the design of the computing circuit used in conformal mapping is presented below.

Given: the z plane in which the point $x+iy$ is to be plotted, and the w plane in which the point $u+iv$ is to be plotted. Let the function $w=w(z)$ define a transformation of the z plane into the w plane.

* Now at International Business Machines Corporation, Boston, Massachusetts.
† T. M. Goodyear Aircraft Corporation, Akron 15, Ohio.
[1] C. A. Meneley and C. D. Morrill, Proc. Inst. Radio Engrs. 41, 1487 (1953).

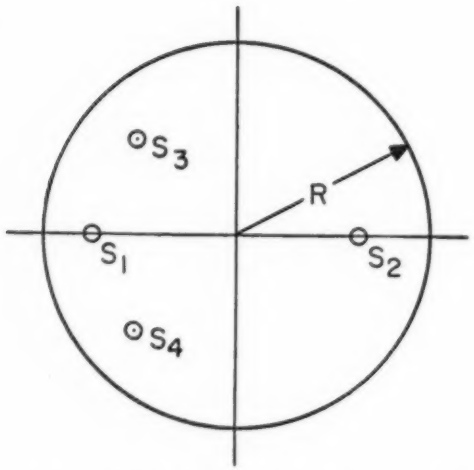


FIG. 1. Location of the roots relative to center of circle.

Whenever a point is given in the z plane, the function $w=w(z)$ determines a value of $u+iv$ that designates a point in the w plane. As z ranges over any configuration in the z plane, the corresponding point $(u+iv)$ describes a figure in the w plane.

For this particular problem, a circle in the z plane is transformed into an airfoil shape in the w plane. The mapping function for this transformation is

$$w=w(z)=z+(a/z)+(b/z^2)+(c/z^3), \tag{1}$$

where a, b, and c are real constants.

Substituting the identities

$$w=u+iv \tag{2}$$

and

$$z=x+iy \tag{3}$$

into Eq. (1) and separating the real and imaginary parts result in:

$$u=x+\frac{ax}{x^2+y^2}+\frac{b(x^2-y^2)}{(x^2+y^2)^2}+\frac{c(x^3-3xy^2)}{(x^2+y^2)^3} \tag{4}$$

and

$$v=y-\frac{ay}{x^2+y^2}-\frac{2bxy}{(x^2+y^2)^2}+\frac{c(y^3-3x^2y)}{(x^2+y^2)^3}. \tag{5}$$

The parametric equations of the circle in the z plane are

$$x=x_0+R\sin\omega t \tag{6}$$

and

$$y=y_0+R\cos\omega t. \tag{7}$$

A variety of airfoil shapes can be obtained through various combinations of the constants in Eqs. (4), (5), (6), and (7). However, if one particular airfoil shape is sought, a judicious selection of constants can be made by suitable arrangement of the roots of

$$(dw/dz)=1-(a/z^2)-(2b/z^3)-(3c/z^4)=0. \tag{8}$$

This procedure reduces the number of trials required to find the correct constants of Eq. (1) for mapping a circle into a specified airfoil shape.

The relations of the constants a, b, and c to the roots z_1, z_2, z_3, and z_4 of Eq. (8) are derived as follows. Equation (8) can be written as

$$z^4-az^2-2bz-3c=0. \tag{9}$$

Moreover,

$$z^4-az^2-2bz-3c=(z-z_1)(z-z_2)(z-z_3)(z-z_4). \tag{10}$$

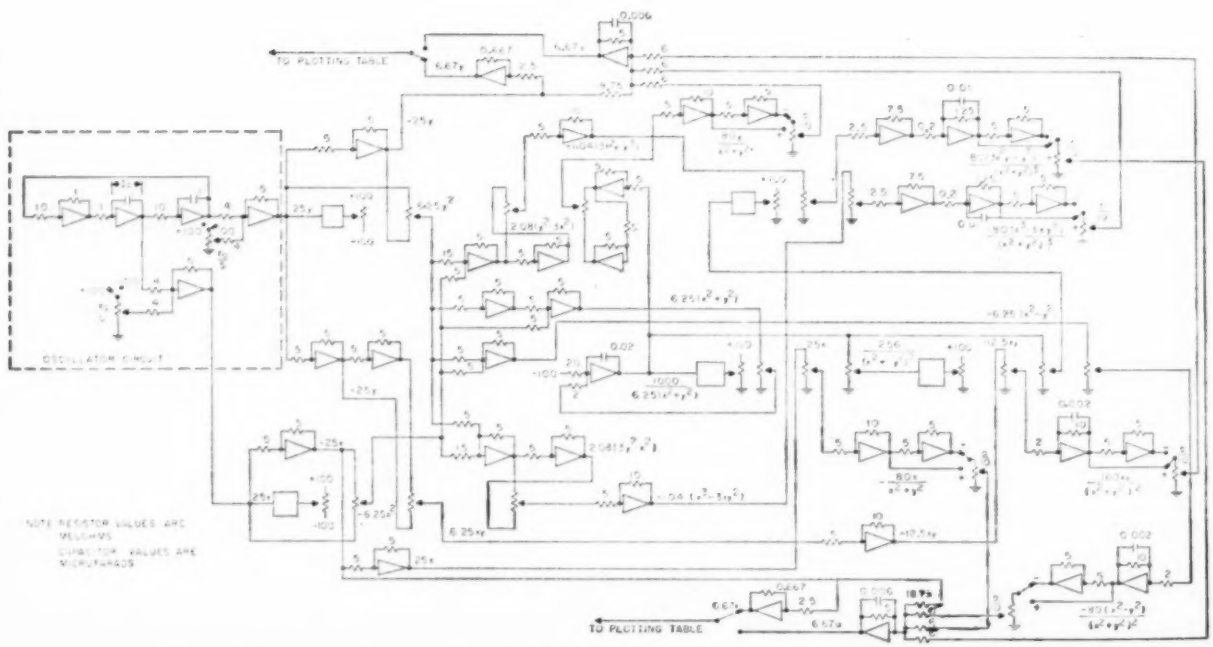


FIG. 2. Wiring diagram of analog-computing circuits employed for construction of conformal maps.

Equating coefficients of Eq. (10) yields

$$0=z_1+z_2+z_3+z_4, \tag{11}$$

$$a=-(z_1z_2+z_1z_3+z_1z_4+z_2z_3+z_2z_4+z_3z_4), \tag{12}$$

$$b=\tfrac{1}{2}[z_1z_2(z_3+z_4)+z_3z_4(z_1+z_2)], \tag{13}$$

and

$$c=-\tfrac{1}{3}(z_1z_2z_3z_4). \tag{14}$$

Let the position of the roots, relative to the center of the circle, be designated s_1, s_2, s_3, and s_4, as shown in Fig. 1. Since the location of the center of the circle in the z plane is x_0+iy_0, then

$$z_1=s_1+x_0+iy_0, \tag{15}$$

$$z_2=s_2+x_0+iy_0, \tag{16}$$

$$z_3=s_3+x_0+iy_0, \tag{17}$$

and

$$z_4=s_4+x_0+iy_0. \tag{18}$$

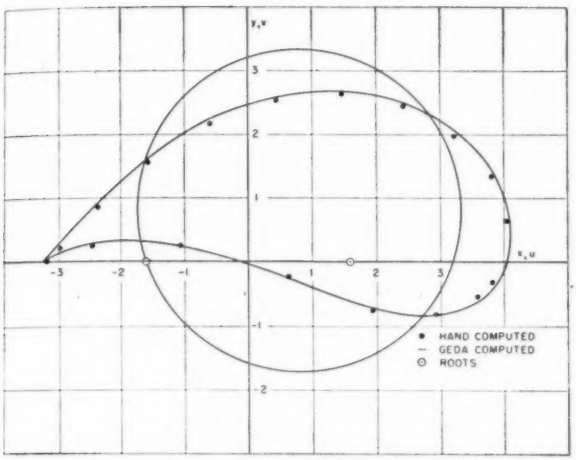


FIG. 3. Comparison of analog-computed and hand-computed asymmetrical airfoil shapes.

Substituting Eqs. (15), (16), (17), and (18) into Eq. (11) results in

$$s_1+s_2+s_3+s_4+4x_0+4iy_0=0. \tag{19}$$

If s_1 and s_2 are real values and s_3 and s_4 are conjugate pairs, then from Eq. (19)

$$x_0=\frac{s_1+s_2+s_3+s_4}{4} \tag{20}$$

and

$$y_0=0. \tag{21}$$

Thus, the location of the roots s_1, s_2, s_3, and s_4 determines the center of the circle and the constants a, b, and c.

It can be seen from Eqs. (4) and (5) that

$$u(x,y)=u(x,-y) \tag{22}$$

and

$$v(x,y)=-v(x,-y). \tag{23}$$

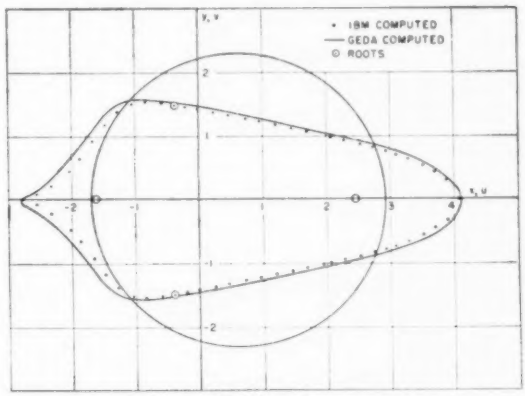


FIG. 4. Comparison of analog-computed and IBM-computed airfoil shapes.

Thus, if the center of the circle is on the x axis, the airfoil shape will be a symmetrical pattern about the u axis.

Stream Function

The complex analytic function

$$F(z)=\phi(x,y)+i\psi(x,y) \tag{24}$$

satisfies Laplace's differential equation

$$\frac{\partial^2 F}{\partial x^2}+\frac{\partial^2 F}{\partial y^2}=0. \tag{25}$$

The velocity potential $\phi(x,y)$ and the stream function $\psi(x,y)$ are real functions. Each satisfies Eq. (24).

The flow about a circle centered at the origin of the z plane in a uniform stream is given as

$$\psi(x,y)=Vy\left(1-\frac{R^2}{x^2+y^2}\right), \tag{26}$$

where V is the free-stream velocity and R is the radius of the circle.[2] The flow past an airfoil can be determined from the flow past a circle, since $\psi(x,y)$ satisfies Laplace's equation, and solutions of the equation remain solutions under a conformal transformation.

Computer Wiring Diagram

The wiring diagram for the analog-computer solution of Eqs. (4) and (5) is shown in Fig. 2. The scale factor throughout the circuit is one volt per unit. The circle is generated by a low-frequency oscillator circuit.

The constants a, b, c, x_0, and y_0 are readily adjusted with potentiometers, and the radius of the circle is controlled by the initial-condition voltage of the oscillator circuit. According to the wiring diagram, a point

[2] Ira H. Abbott and Albert E. von Doenhoff, *Theory of Wing Sections* (McGraw-Hill Book Company, Inc., New York, 1949), p. 41.

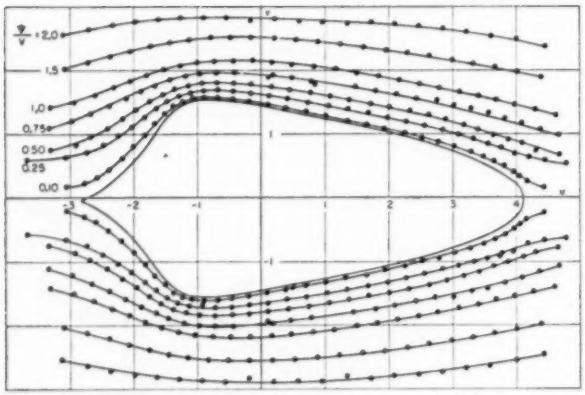


FIG. 5. Analog-computed airfoil shape and streamlines.

in the z plane is restricted so that

$$1.6<(x^2+y^2)<16. \tag{27}$$

Therefore, the maximum analog voltage of 100 is not exceeded.

RESULTS

Figures 3, 4, and 5 show representative airfoil shapes constructed in the analog-computing equipment. The figures were resolved on an $X-Y$ plotting board.

Accuracy of analog computation is comparable to that of desk calculation, as shown in Fig. 3. Figure 4 is a comparison of Geda and IBM solutions. The streamlines about an airfoil are illustrated in Fig. 5. The streamlines were transformed point-by-point (using analog-computing equipment) into the w plane from the known streamlines about a circle (Eq. (26)).

TABLE I. Analog-computing input information used in determining airfoil shapes.

Input	Value Fig. 3	Value Fig. 4	Value Fig. 5
a	2.56	2.18	2.18
b	0.000	2.47	2.47
c	0.000	3.05	3.05
x_0	0.800	0.638	0.638
y_0	0.800	0.000	0.000
R	2.53	2.30	2.30
z_1	−1.60	−1.62	−1.62
z_2	1.60	2.41	2.41
z_3	not defined	$-0.395+i\times1.48$	$-0.395+i\times1.48$
z_4	not defined	$-0.395-i\times1.48$	$-0.395-i\times1.48$

Table I gives the values of the constants, position and radius of the circle, and the position of the roots for each figure.

DISCUSSION

The advantages of this method of conformal mapping are the speed and ease of computation. Table II summarizes the man-hours required for three mapping methods: (1) analog computation; (2) card-programmed calculation; and (3) manual computation using a desk calculator.

It will be noted from Table II that for a problem requiring the construction of seven or more shapes the analog computer offers considerable advantage. In many exploratory studies, the number of shapes required may be much greater than seven.

TABLE II. Comparison of time required to set up and plot conformal maps with three types of computers.

Type of computer	Problem setup time	Construction time/ one shape
Analog computer	16 hr	5 min
Card-programmed calculator	10 hr	1 hr
Manual computist	negligible	20 hr

It is apparent that the use of an analog computer enables one to explore quickly and accurately a great variety of airfoil shapes and their streamlines. Since an analog computer requires only a relatively short time for solution and saves a considerable amount of numerical computation, many boundary-value problems of applied mathematics in fields other than fluid dynamics may be treated by this method. Recent studies[3–6] have investigated a variety of applications of conformal mapping. In such applications the use of analog computers may be highly advantageous.

[3] H. Poritsky, "Some industrial applications of conformal mapping," Proceedings of a Symposium on Construction and Applications of Conformal Maps, Natl. Bur. of Standards (Government Printing Office, Washington, D. C., December 26, 1952), pp. 15 and 18.

[4] G. M. Stein, "Conformal maps of electric and magnetic fields in transformers and similar apparatus," reference 3, pp. 31 and 32.

[5] Ernst Weber, "Conformal mapping applied to electromagnetic field problems," reference 3, p. 64.

[6] Proceedings of the First U. S. Natl. Congr. Appl. Mechanics (Am. Soc. Mech. Engrs., New York, 1952), p. 227.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 2 FEBRUARY, 1955

On the Diffusion of Decaying Particles

LOUIS GOLD
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received March 15, 1954)

The probability of neutron absorption on a spherical surface for a radial source in a medium for which the particle has prescribed lifetime and rate of diffusion is calculated. This helps illuminate the more intricate, related problem for a charged particle diffusing in the presence of a radial inverse-square electric field to the extent that an explicit relation for the relative absorption probabilities that accompany changes in the direction of the drift field is secured. An asymptotic solution for the absorption probability is also provided which in first-order approximation may be applied in the range of interest for transistors.

THIS communication is the outcome of an investigation to apply the original findings of J. Keilson[1] on the diffusion of decaying particles in a radial electric field (appropriate to hole diffusion in semiconductors) to the related neutron diffusion problem. The later, in effect, represents the reduced situation of the field-free case or the zero charged particle and so allows explicit solution where otherwise some rather involved approximational schemes and numerical solutions were offered in the prototype study.

Now while the neutron case is somewhat elementary, it does help illuminate some of the more obscure points in the (JK) paper and leads to findings not contained in this earlier work. In fact, as will be demonstrated, one can relate, very simply, the effect of an attractive or repulsive field on the probability of a charged particle being absorbed at a collector electrode. Moreover, an asymptotic solution can be developed which allows for explicit characterization of this probability—without recourse to numerical or graphical solutions.

I. THE NEUTRON PROBLEM AS A SPECIAL CASE IN THE DIFFUSION OF DECAYING PARTICLES

The problem is of the following general nature. Particles enter an infinite medium at some point and then proceed to diffuse in the presence of a radial inverse-square intensity electric field which has its center at another point about which is a hemispherical surface of radius r_0 that absorbs all incident particles. If a prescribed number of particles per second are injected into the medium, how many per second will be absorbed before decaying by some such process as recombination.

As pointed out by (JK), it is difficult to treat the problem in any general manner, but if one deals with a radial current source and formulates a one-dimensional, radial flow, the probability of absorption γ can be calculated since all points on the radial source have the same γ. Thus the basic differential equation describing the steady state flow in terms of the reduced distance s may be expressed in the form (neglecting space charge)

$$\frac{d^2\rho}{ds^2}+\left(\frac{2}{s}\pm\frac{\lambda}{s^2}\right)\frac{d\rho}{ds}-\rho=0\begin{cases}\lambda=\mu\epsilon_0(\tau/D)^{\frac{1}{2}}\\ s=r/(D\tau)^{\frac{1}{2}}\end{cases}\tag{1}$$

where ρ is the particle density, D the diffusion coefficient, τ the lifetime, and λ a dimensionless parameter which also includes the mobility μ and the applied field parameter as defined by conventional radial electric field $\mathbf{\varepsilon}$ relation $\mathbf{\varepsilon}=(\epsilon_0/r^2)\mathbf{u}_r$, where $\mathbf{u}_r$ is unit vector. The minus sign denotes the repulsive case while the positive sign holds for the attractive case, both being associated with the drift contribution to the diffusion.

The problem may be visualized in terms of Fig. 1 which shows the gross appearance of $\rho(s)$ with s_0 and a_0, the respective locations of the absorbing surface and the source site. The boundary conditions that prevail are

$$\begin{aligned}&\rho_A(s_0)=0 \quad \rho_A'(a_0)\neq\rho_B'(a_0)\\ &\rho_B(\infty)=0 \quad \rho_A(a_0)=\rho_B(a_0)\end{aligned}\tag{2}$$

from which is to be found γ, the probability that a particle introduced at a_0 will arrive at the absorption point s_0, *viz.*,

$$\gamma=\frac{s_0^2\rho_A'(s_0)}{a_0^2[\rho_A'(a_0)-\rho_B'(a_0)]}.\tag{3}$$

For the neutron diffusion where $\lambda=0$, the general solution of (1) can be verified to be

$$\rho=A'\frac{\cosh s}{s}+B'\frac{\sinh s}{s}$$

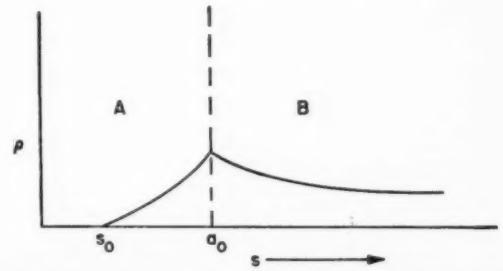


FIG. 1. Stationary particle density distribution about source a_0. The absorbing surface located at s_0 beyond which $\rho=0$. In region B, $\rho\to 0$ as $s\to\infty$.

[1] J. Keilson, J. Appl. Phys. **24**, 1397 (1953)—hereafter referred to as (JK).

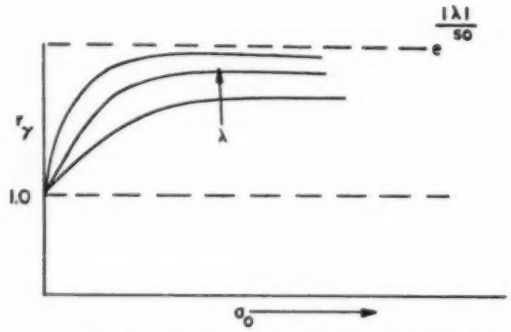


FIG. 2. Relative change in the absorption probability for electric field reversal. Direction of arrow for λ shows enhanced effect as the magnitude of the field increases with increasing sensitivity as a_0 increases until the asymptotic value $e^{|\lambda|/s_0}$ is attained.

but for convenience we write this in the equivalent form

$$\rho=\frac{A}{s}e^{-s}+\frac{B}{s}e^{s}. \tag{5}$$

Thus, in accord with the dictates of Fig. 1 and its boundary requirements, it follows that

$$\begin{cases}\rho_A=\dfrac{A}{s}(e^{-s}-e^{s-2s_0})\\[2ex] \rho_B=\dfrac{A}{s}[1-e^{2(a_0-s_0)}]e^{-s}\end{cases} \tag{6}$$

where the constant A is a measure of the source intensity.

The probability coefficient may be found directly as prescribed by (3). However, it is more efficient to make use of the transformation to

$$\begin{cases}g''(s)=g(s)\\ \rho(s)=\dfrac{1}{s}g(s),\end{cases} \tag{7}$$

which allows for a labor-saving expression

$$\gamma=\frac{s_0}{a_0}\frac{g_A'(s_0)}{g_A'(a_0)-g_B'(a_0)} \tag{8}$$

because of the boundary conditions imposed. Here we have

$$\begin{cases}g_A'=-A(e^{-s}+e^{s-2s_0})\\ g_B'=-A[1-e^{2(a_0-s_0)}]e^{-s}\end{cases} \tag{9}$$

whereby using (8) yields

$$\gamma=\frac{s_0}{a_0}e^{-(a_0-s_0)}\qquad a_0\geq s_0. \tag{10}$$

Clearly this result for the probability coefficient can be secured directly from the suitably chosen solution of the differential equation for γ adjoint to (1) with the appropriate boundary conditions $\gamma(s_0)=1$ and $\gamma(\infty)=0$.

II. SOME GENERAL PROPERTIES WHERE $\lambda\neq 0$

A proper understanding of the (JK) analysis allows for ready deduction of the relation between the repulsive and attractive probability coefficients. Thus if one poses the change of variable

$$\rho(s)=e^{\pm\lambda/2s}\frac{g(s)}{s} \tag{11}$$

where the signs are chosen to correspond to (1) then the outcome

$$g''=(1+\lambda^2/4s^2)g \tag{12}$$

is independent of sign. Consequently for the derived

$$\gamma=\frac{s_0}{a_0}\frac{e^{\pm\lambda/2a_0}\,g_1(a_0)}{e^{\pm\lambda/2a_0}\,g_1(s_0)} \tag{13}$$

g_1, which represents the negative exponential solution of (12), is unaffected by sign considerations with the outcome

$$r_\gamma=\frac{\gamma\text{ attractive}}{\gamma\text{ repulsive}}=\exp\left[|\lambda|\left(\frac{1}{s_0}-\frac{1}{a_0}\right)\right],\quad a_0\geq s_0. \tag{14}$$

The behavior of r_γ is depicted in Fig. 2 which shows the ever-widening gap between the attractive and repulsive probability coefficient as $a_0\to\infty$ with an upper limiting value of $e^{|\lambda|/s_0}$.

III. THE ASYMPTOTIC SOLUTION FOR THE PROBABILITY COEFFICIENT

We next present an asymptotic solution of Eq. (12) which allows ready characterization of γ without resort to graphical or numerical methods. Since, as stated in relation (13), only the negative exponential solution g_1 of the differential Eq. (12) is required for the problem at hand, the presumed form of the asymptotic solution is taken to be

$$g_1=e^{-s}\{s^\nu+a_1s^{\nu-1}+a_2s^{\nu-2}+a_3s^{\nu-3}+\cdots\} \tag{15}$$

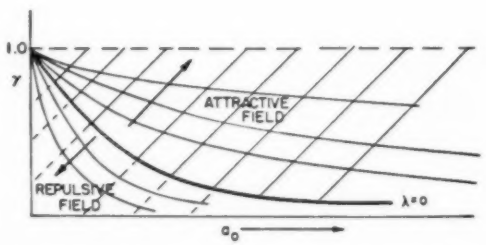


FIG. 3. General behavior of the absorption probability as the separation of source and collector increases. The arrows in the fields on either side of $\lambda=0$ show the direction of $|\lambda|\to\infty$.

whence substitution into (12) yields

$$2\{\ \}_{II}+\{\ \}_{III}-\lambda^2/4s^4\{\ \}_I=0 \tag{16}$$

wherein we have employed the scheme

$$\{\ \}_{II}=\nu s^{\nu-1}+a_1(\nu-1)s^{\nu-2}+a_2(\nu-2)s^{\nu-3}+\cdots$$
$$\{\ \}_{III}=\nu(\nu-1)s^{\nu-2}+a_1(\nu-1)(\nu-2)s^{\nu-3}$$
$$+a_2(\nu-2)(\nu-3)s^{\nu-4}+\cdots \tag{17}$$

and $\{\ \}_I$ is the bracket factor in (15).

Then we find outright

$$\nu=0,\quad a_1=a_2=0$$
$$a_3=\frac{\lambda^2}{6\cdot 4},\quad a_4=\frac{-12}{8}\cdot\frac{\lambda^2}{6\cdot 4},\quad a_5=\frac{5\cdot 4}{5\cdot 2}\frac{\lambda^2}{16} \tag{18}$$

with the general recursion formula for $n\geq 5$

$$2(n-1)a_{n-1}+(n-2)(n-1)a_{n-2}-\frac{\lambda^2}{4}a_{n-4}=0 \tag{19}$$

so that

$$a_6=\frac{-\lambda^2}{8\cdot 6\cdot 2}(30-\lambda^2/12)$$

reveals the involved way in which λ enters. Hence our sought-for relation becomes

$$g_1=e^{-s}\left[1+\frac{\lambda^2}{4s^3}S(s)\right] \tag{20}$$

in which the function $S(s)$ represents the series

$$S(s)=\frac{1}{6}-\frac{1}{4s}+\frac{1}{2s^2}-\frac{1}{24}\left(30-\frac{\lambda^2}{12}\right)\frac{1}{s^3}+\cdots \tag{21}$$

We establish then the central expression as

$$\gamma=\frac{s_0}{a_0}e^{-(a_0-s_0)}\cdot\frac{e^{\pm\lambda/2s_0}}{e^{\pm\lambda/2a_0}}\cdot\frac{1+\dfrac{\lambda^2}{4a_0^3}S(a_0)}{1+\dfrac{\lambda^2}{4s_0^3}S(s_0)} \tag{22}$$

which correctly reduces to (10), as it should, for $\lambda=0$.

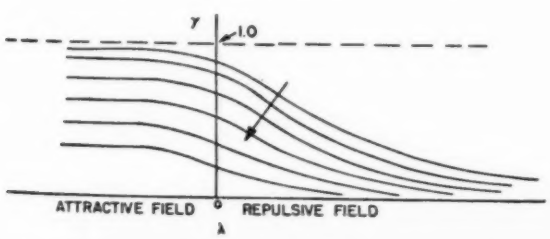


FIG. 4. Schematic presentation of absorption probability variation with λ. Direction of arrow indicates increasing parameter a_0.

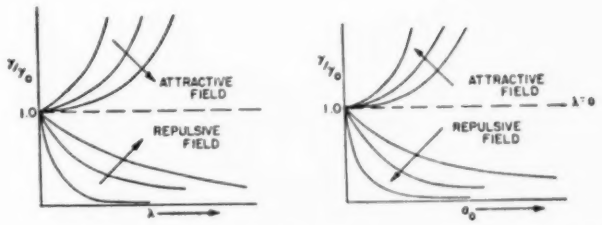


FIG. 5(a). Comparative absorption probabilities for field dependent and field independent diffusion. The direction of the arrow in the two fields indicates influence of parameter $a_0\to\infty$. (b). Comparative absorption probabilities *versus* source-collector separation with λ the parameter increasing along the directions of the arrows in the attractive and repulsive fields.

IV. FIRST-ORDER APPROXIMATION AND THE CHARACTERIZATION OF γ

Clearly, our γ relation (22) cannot be employed with impunity for all values of λ, a_0 and s_0. The series (21) will converge nicely for the range of λ and s indicated

$$0\leftarrow\lambda\leq 1$$
$$\infty\leftarrow s\geq 1, \tag{23}$$

but will be in difficulty otherwise. We confine our attention to the appropriate range where (22) is serviceable; the treatment of hole diffusion has $|\lambda|$ usually within the range of (23) so the asymptotic solution can be helpful.

We write several equivalent forms of λ which result from the well-known Einstein relation $\mu=(q/kT)D$, *viz.*,

$$\lambda=\frac{q\epsilon_0}{kT}(D\tau)^{\frac{1}{2}}=\epsilon_0\left(\frac{\mu\tau q}{kT}\right)^{\frac{1}{2}}. \tag{24}$$

Whence at 300°K and mobilities of the order of 10^3 cm^2/volt-sec and lifetimes in the domain of microseconds λ is reasonable for the application of a first order approximation when ϵ_0 is $\sim$1 volt-cm. Accordingly we employ

$$\gamma=\gamma_0\exp\left[\pm\frac{\lambda}{2}\left(\frac{1}{s_0}-\frac{1}{a_0}\right)\right], \tag{25}$$

with γ_0 the probability coefficient for $\lambda=0$ as a replacement for (JK) scheme to circumscribe γ in the range of $|\lambda|$ from 1 to 0.1.

In Fig. 3 is illustrated the general appearance of $\gamma(a_0)$ for prescribed s_0 with λ as the parameter. The curve for $\lambda=0$ splits the field into the attractive and repulsive domains, the former moving toward $\gamma(a_0)=1$ and the latter toward $a_0=s_0$ as $|\lambda|\to\infty$. The companion Fig. 4 exhibits the behavior of $\gamma(\lambda)_{a_0}$.

The next set of Fig. 5a and 5b shows schematically how γ/γ_0 varies respectively with λ and a_0 so that hist essentially completes the delineation of the probability coefficient, explicitly.

On Power Spectra and the Minimum Detectable Signal in Measurement Systems

J. J. FREEMAN
National Bureau of Standards, Washington, D. C.
(Received February 1, 1954)

A least upper bound for the minimum detectable signal of a receiving system whose input is a periodic wave form plus interference is specified in terms of the first and second moments of the rectified output. Also, it is shown that the power spectrum of the rectified output (together with the impedance characteristic of the output meter) serves to establish the first and second moments of the meter deflection for either of the following two cases:

(1) The rectified output operates a meter, and the difference in meter readings with the signal on and off constitutes the observation from which the message is inferred.

(2) The rectified output feeds a synchronous detector whose meter reading is directly proportional to the message. Here we require that there be no input noise which is modulated at the synchronous frequency.

I. INTRODUCTION

THIS paper demonstrates two theorems. First, a least upper bound for the minimum detectable signal of any measurement system whose input is a periodic wave form plus interference may be specified by the first and second moments of the rectified signal. Second, the power spectrum of the rectified output (together with the impedance characteristics of the output meter) serves to establish the first and second moments of the meter deflection for either of the following two cases:

(1) The rectified output operates a meter, and the difference in meter readings with the signal on and off constitutes the observation from which the message is inferred.

(2) The rectified output feeds a synchronous detector whose meter reading is directly proportional to the message. Here we require that there be no input noise which is modulated at the synchronous frequency.

Although the foregoing theorems are elementary, it is believed they are sufficiently fundamental to warrant explicit demonstration.

II. MINIMUM DETECTABLE SIGNAL

In most discussions of detectability criteria, it is postulated that the threshold of detectability occurs when the detected noise power approximates the signal power.[1,2] Other more complete discussions[3] establish the detectability criteria in terms of the probability distribution of the detected signal. Unfortunately, this probability distribution is often unavailable in actual nonlinear systems. However, the first and second moments of the detected signal are available, either theoretically or experimentally, and this section will establish a least upper bound for the minimum detectable signal in terms of the first two moments of the observed or rectified signal.

Consider a receiver whose input consists of interference plus a periodic wave containing a definite message (such as a constant modulation parameter, or peak intensity) which may be characterized by a single number. The interference is assumed to consist of noise $u(t)$, (a stationary random wave form), unwanted periodic wave forms, periodically modulated noise, or a sum of these. The receiver will extract the message by means of a proper demodulating element which will operate a meter. The apparatus is turned on at $t=0$, and after a lapse of time sufficient for transients to die out the observer will record the meter reading, $M(t_1)$ at an arbitrary time, t_1, thereafter. More generally the observer might choose to read an average over an interval τ, namely,

$$\frac{1}{\tau}\int_{t_1}^{t_1+\tau} M(t)dt, \tag{1}$$

or an average of two readings, namely,

$$\frac{M(t_1)+M(t_1+\tau)}{2}.$$

However, the observation of these functions of $M(t)$ is equivalent to merely modifying the transfer characteristic $Z(f)$ of the meter. Thus, reading the foregoing continuous average is equivalent to multiplying Z by the factor $(e^{j\omega\tau}-1)/j\omega\tau$, and reading the above discrete average is equivalent to multiplying Z by $(1+e^{j\omega\tau})/2$, where $\omega=2\pi f$. Thus, no loss in generality results from considering only the case where the meter is observed at a single instant.

The presence of interference, which includes noise, will cause the observed meter reading to be a random variable. In the absence of any *a priori* knowledge an observation at any instant is assumed to be equivalent to an observation at any other. Thus, a particular reading depends (with uniform probability) upon the

[1] J. Lawson and G. Uhlenbeck, *Threshold Signals* (McGraw-Hill Book Company, Inc., New York, 1950), p. 161.
[2] C. Eckart, "The Measurement and Detector of Steady ac and dc Signals in Noise," Report 51–39, Oct. 4, 1951, Marine Physical Laboratory of the Scripps Institution of Oceanography.
[3] D. Middleton, "Statistical Criteria for the Detection of Pulsed Carriers in Noise," Report E5091, November, 1952, Air Force Cambridge Research Center, Cambridge, Massachusetts.

instant of observation as well as upon the particular member of the noise ensemble. Explicitly, $M(t)=M(u(t),s,t)$ where s is the message parameter, $u(t)$ the input noise. The functional dependence of the meter reading on noise, periodic interference, and message parameter is assumed to be known. The mean and the variance of $M(t)$ will be definite numbers which are assumed to be known functions of s, a definite number.

$$\langle M(t)\rangle=\bar{M}(s)=g(s) \tag{2a}$$

and

$$\langle(M(t)-\bar{M})^2\rangle=\sigma^2(s)=h(s), \tag{2b}$$

where the angular brackets indicate the average is taken over all random variables, and $\bar{M}(s)$ signifies the average meter reading corresponding to the particular value s of message parameter. It is assumed that Eq. (2a) may be solved for s. Thus,

$$s=G(\bar{M}). \tag{2c}$$

The observer, however, will use $M(t_1)$ not as the outcome of a single random experiment but rather as a causal event, and will infer a value of s by substituting $M(t_1)$, Eq. (2c), namely

$$s_1=G(M(t_1)). \tag{3}$$

The question arises concerning the smallest value of message parameter which is detectable. By this one means[4] the smallest value of s, such that, if $s>0$ (the signal exists), this will be observed with probability $>\alpha$, a prescribed constant; also, if no signal exists ($s=0$), *that* will be observed with probability $>\beta$ (also prescribed). A signal is said to be observed ($s>0$) if $M(t)>M'$, to be determined. No signal is said to be observed ($s=0$) if $M(t)<M'$.

We require, then, the smallest value of s such that

$$P[M(t)>M'|s>0]>\alpha \tag{4}$$

and

$$P[M(t)<M'|s=0]>\beta, \tag{5}$$

where $P[M(t)>M'|s>0]$ reads "the probability that $M(t)>M'$, given that $s>0$."

A least upper bound to this minimum detectable s is given by Tshebysheff's one-sided inequality[5] which asserts that if X be a stochastic variable with mean$=0$, and standard deviation σ, then

$$P(X\leqslant\chi)\geqslant\chi^2/(\sigma^2+\chi^2)\quad\chi>0 \tag{6}$$

$$P(X\geqslant\chi)\geqslant\chi^2/(\sigma^2+\chi^2)\quad\chi<0. \tag{7}$$

[4] D. O. North, "An Analysis of the Factors which Determine Signal/Noise Discrimination in Pulsed/Carrier Systems," R.C.A. Report, June 25, 1943.

[5] Uspensky, *Introduction to Mathematical Probability, Problem 1* (McGraw-Hill Book Company, Inc., New York, 1937), first edition, p. 198.

Choose M' such that $\bar{M}(0)<M'<\bar{M}(s)$. Then from Eq. (7),

$$P[M-\bar{M}(s)>M'-\bar{M}(s)|s>0]=P[M>M'|s>0]\geqslant\frac{(M'-\bar{M}(s))^2}{\sigma^2(s)+(M'-\bar{M}(s))^2}. \tag{8}$$

From Eq. (6),

$$P[M-\bar{M}(0)<M'-\bar{M}(0)|s=0]=P[M<M'|s=0]\geqslant\frac{(M'-\bar{M}(0))^2}{\sigma^2(0)+(M'-\bar{M}(0))^2}. \tag{9}$$

Hence if

$$\alpha=\frac{(M'-\bar{M}(s))^2}{\sigma^2(s)+(M'-\bar{M}(s))^2},\ \text{Eq. (4) is satisfied.} \tag{10}$$

And if

$$\beta=\frac{(M'-\bar{M}(0))^2}{\sigma^2(0)+[M'-\bar{M}(0)]^2},\ \text{Eq. (5) is satisfied.} \tag{11}$$

M' is determined from Eq. (11), namely

$$M'=\bar{M}(0)+\sigma(0)(\beta/1-\beta)^{\frac{1}{2}}. \tag{12}$$

This value of M' substituted in Eq. (10) yields a lower bound for the minimum detectable meter reading and its associated message parameter, namely,

$$\bar{M}(s)=\bar{M}(0)+\sigma(0)(\beta/1-\beta)^{\frac{1}{2}}+\sigma(s)(\alpha/1-\alpha)^{\frac{1}{2}}. \tag{13}$$

Designate the value of s satisfying Eq. (13) by s'. Then any $s\geqslant s'$ is detectable.* For if $s\geqslant s'$, $M(s)\geqslant\bar{M}(s')$, and $\bar{M}(s)/\bar{M}(s')>\sigma^2(s)/\sigma^2(s')$, so that

$$P[M>M'|s\geqslant s']\geqslant\frac{(\bar{M}(s)-M')^2}{\sigma^2(s)+(\bar{M}(s)-M')^2}\geqslant\frac{(\bar{M}(s')-M')^2}{\sigma^2(s')+(\bar{M}(s')-M')^2}\geqslant\alpha. \tag{14}$$

Hence, all values of $s\geqslant s'$ are detectable.

If the probability distributions of $M(t)$ were known, Eqs. (4) and (5) would determine the minimum detectable signal precisely. However, Eq. (13) gives the best[6] estimate of the minimum detectable signal in terms of the first and second moments of the probability distribution, with prescribed α and β.

III. POWER SPECTRA IN MEASUREMENT SYSTEMS

In this section we demonstrate the utility of the concept of a power spectrum applied to a nonlinear measurement system in which the input consists of periodic wave forms plus noise. The input here is not stationary owing to the presence of periodic wave forms. However, the output observation is a stationary random process since the instant at which an observation is

* We assume that the first moment increases faster than the second moment with increasing m.

[6] Uspensky, see reference 5.

made (or sequence of observations begins) is considered a random variable, so that the statistics of any observation or sequence of observations is independent of the origin of the time axis.

For example, the measurement system might consist of a cw generator which feeds through a gas sample whose attenuation is varied periodically, thus amplitude-modulating the cw wave. Following the cell is a rf receiver, including a detecting element. The message parameter here is the modulation factor which gives information regarding the absorptivity of the gas. In this experiment, interference is caused by random noise as well as by distortion or undesired periodic wave forms generated by the measurement system itself. We show below that the power spectrum of the detector output enables one to compute the first and second moments of the meter indication when a meter is connected to the detecting element (Case 1), or when a synchronous detector plus meter is connected to the detecting element (Case 2).

Case 1—Meter Output

Consider a meter of transfer impedance $Z(f)$ connected as output to a receiving system whose input consists of periodic wave forms and noise. Since the message parameter must be extracted from the desired periodic wave form to give a monotonic-meter correspondence, the receiver output, $I(t)$ which is a known function of the input and is nonstationary must contain a dc component. Of course $I(t)$ depends upon the message parameter s. Taking a particular sample of $I(t)$ and assuming $I(t)=0$, for $|t|>T/2$ we represent $I(t)$ within the interval $-T/2<t<T/2$ as follows:

$$I(t)=\int_{-\infty}^{\infty}F(f)e^{j\omega t}df=1/T\int_{-T/2}^{T/2}I(t)dt+\int_{-\infty}^{\infty}F'(f)e^{j\omega t}df.$$

Denote

$$1/T\int_{-T/2}^{T/2}I(t)dt$$

as $\bar{I}$, the dc component. Then

$$F'(f)=\int_{-T/2}^{T/2}(I(t)-\bar{I})e^{-j\omega t}dt=F(f)-\bar{I}\frac{\sin\omega T/2}{\omega/2}, \quad (15)$$

where $F'(0)=0$, and $F'(f)$ will be continuous in the neighborhood of $f=0$ for finite T.

The meter reading corresponding to this particular sample of $I(t)$ at a random instant will be

$$M(t)=\int F(f)Ze^{j\omega t}df.$$

Since a meter reading may be taken with probability density dt/T anywhere over the interval $-T/2\leqslant t\leqslant T/2$ the average meter reading, averaged over the noise ensemble as well as the time, is

$$\langle M(t)\rangle_{Av}=1/T\int_{-\infty}^{\infty}\int_{-T/2}^{T/2}\bar{F}(f)Z(f)e^{j\omega t}dfdt$$
$$=\int_{-\infty}^{\infty}\bar{F}(f)Z(f)\frac{\sin\omega T/2}{\omega T/2}df.$$

Letting $\omega T/2=\theta$,

$$\lim_{T\to\infty}\bar{M}=\lim_{T\to\infty}\frac{1}{\pi T}\int_{-\infty}^{\infty}\bar{F}(\theta/\pi T)Z(\theta/\pi T)\frac{\sin\theta}{\theta}d\theta$$
$$=\lim_{T\to\infty}Z(0)\frac{\langle F(0)\rangle_{Av}}{T}. \quad (16)$$

The zero point of the meter, M_0, must be fixed either theoretically or experimentally by determining the average deflection for no modulation and will be considered as a known parameter of the system.

$$\bar{M}_0=\lim_{T\to\infty}Z(0)\bar{F}_0(0)/T,$$

where the subscript in $\bar{F}_0(0)$ designates $s=0$.

$$\langle\bar{M}^2\rangle_{Av}=1/T\int_{-T/2}^{T/2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\langle F(f)F(f')\rangle_{Av}$$
$$\times Z(f)Z(f')e^{j(\omega+\omega')t}dfdf'dt.$$

Similarly,

$$\lim_{T\to\infty}\bar{M}^2=\lim_{T\to\infty}1/T\int_{-\infty}^{\infty}\langle|F(f)|^2\rangle_{Av}|Z|^2df$$
$$=\int_{0}^{\infty}P(f)|Z(f)|^2df, \quad (17)$$

where $P(f)=\lim_{T\to\infty}2/T\langle|F(f)|^2\rangle_{Av}$ is defined as the power spectrum of the receiver output.

For any sufficiently small, but nonzero ϵ,

$$\lim_{T\to\infty}2/T\int_{0}^{\epsilon}\langle|F(f)|\rangle_{Av}^2|Z|^2df=\lim_{T\to\infty}\frac{2Z^2(0)}{T}$$
$$\times\int_{0}^{\epsilon}\left\langle\left[F'(f)+\bar{I}\frac{\sin\omega T/2}{\omega/2}\right]^2\right\rangle_{Av}df=\bar{I}^2Z^2(0)=\bar{M}^2 \quad (18)$$

since $F'(0)=0$.

Subtracting Eq. (18) from Eq. (17),

$$\langle(M-\bar{M})^2\rangle_{Av}=\sigma_M{}^2=\int_{\epsilon}^{\infty}P(f)|Z(f))|^2df. \quad (19)$$

We define "signal" as $\langle\theta(t)\rangle_{Av}=\langle M(t)\rangle_{Av}-M_0$, the average deflection of the meter caused by the message, and we define "noise" as $\sigma_\theta{}^2$. Since $\sigma_\theta{}^2=\sigma_M{}^2$, Eqs. 18 and 19 show that the signal is obtained by integrating the power spectrum of the output (weighted by the meter gain characteristic) over an infinitesimal frequency range in the neighborhood of the origin, the "noise" is obtained by integrating the same function

over the entire frequency range, except for the neighborhood of the origin.

Case 2. Synchronous Detector Output

Assume the receiver output is a voltage of the form

$$e(t)=\sum_{j=0} A_j \cos(\omega_j t+\psi_j)+u(t)\sum_k B_k \cos\omega_k' t, \quad (20)$$

where A_0 is the parameter to be measured, the remaining terms of $e(t)$ being interference. We assume $u(t)$ is a stationary random process with a continuous power spectrum.

The receiver is connected to a synchronous detector,[7] one variant of which is illustrated in Fig. 1.

The grids of tubes T_1 and T_2, which are assumed to be operating linearly are, connected in parallel with the input voltage, $e(t)$. (Fig. 1.) The screens of T_1 and T_2 are each connected to the secondary of a balanced transformer whose primary is connected to a source of voltage, $B\cos(\omega_0 t+\psi)$ at the frequency of the desired signal. A phase-shifter (not shown) permits ψ to be varied. The plate currents of T_1 and T_2 are coupled into two balanced thermocouples whose emf's are connected in series with a meter M so that the emf's are opposing. Assuming for simplicity that the time constant of each thermocouple is zero, the emf generated by the first couple will be

$$[ae(t)+b\cos(\omega_0 t+\psi)]^2.$$

The emf generated by the second couple will be

$$[ae(t)-b\cos(\omega_0 t+\psi)]^2.$$

The resulting emf acting on the meter will be the difference, namely,

$$e_m=De(t)\cos(\omega_0 t+\psi), \quad (21)$$

where D is a constant of the system.

Analyzing a sample of $e(t)$ within the interval $-T/2\leqslant t\leqslant T/2$,

$$e=\int_{-\infty}^{\infty} F(f)e^{j\omega t}df, \quad (22)$$

where

$$F(f)=\int_{-T/2}^{T/2} e(t)e^{-j\omega t}dt. \quad (23)$$

From Eq. (21),

$$e_m(t)=D/2\int_{-\infty}^{\infty}\{F(f-f_0)e^{i\psi}+F(f+f_0)e^{-i\psi}\}e^{j\omega t}df. \quad (24)$$

Let

$$H(f)=F(f-f_0)e^{i\psi}+F(f+f_0)e^{-i\psi}. \quad (25)$$

Then if Z is the transfer impedance of the meter (deflection per volt),

$$M(t)=D/2\int_{-\infty}^{\infty} Z(f))H(f)e^{j\omega t}df,$$

where $M(t)$ represents a meter reading at a random instant, t, within the interval $(-T/2\leqslant t\leqslant T/2)$.

$$\bar{M}(t)=\lim_{T\to\infty} D/2\int_{-T/2}^{T/2}\int_{-\infty}^{\infty} Z(f)\bar{H}(f)/Te^{j\omega t}dfdt$$

$$=\lim_{T\to\infty} D/2Z(0)[\bar{F}(-f_0)e^{i\psi}+\bar{F}(f_0)e^{-i\psi}]/T. \quad (26)$$

Substituting Eq. (20) in Eq. (23),

$$\bar{M}(t)=D/2Z(0)A_0\cos(\psi-\psi_0). \quad (27)$$

If ψ is varied until the meter deflection is a maximum and the system constants $DZ(0)$ are known, A_0 can be correctly inferred from $\langle M(t)\rangle_{Av}$. A measurement of A_0 by this method implies a knowledge of the frequency and the phase angle (which is measured by the adjustment for maximum response). We assume henceforth that $\psi_0=\psi=0$ without loss of generality.

Applying the same reasoning outlined in Case 1,

$$\bar{M}^2=\lim_{T\to\infty} D^2/2\int_0^{\infty}|Z|^2\langle|H(f)|^2\rangle_{Av}/Tdf. \quad (28)$$

In practice, the pass band of the meter will be much narrower than the synchronous frequency $[Z(f)=0, |f|>f_0]$ so that Eq. (28) becomes, on substituting the explicit value of H given by Eq. (25),

$$\langle M^2\rangle_{Av}=\lim_{T\to\infty} D^2/2\int_0^{f_0}|Z|^2\{\langle F(f-f_0)F(-f-f_0)\rangle_{Av}/T$$

$$+\langle|F(f-f_0)|^2\rangle_{Av}/T+\langle|F(f+f_0)|^2\rangle_{Av}/T$$

$$+\langle F(f+f_0)F(-f+f_0)\rangle_{Av}/T\}df. \quad (29)$$

$$\langle M^2\rangle_{Av}=\lim_{T\to\infty} D^2/2\int_0^{2f_0}|Z(f-f_0)|^2\langle|F(f)|^2\rangle_{Av}/Tdf$$

$$+\lim_{T\to\infty}\int_0^{f_0}|Z(f-f_0)|^2\langle F(-f)F(f-2f_0)\rangle_{Av}/Tdf$$

$$+\lim_{T\to\infty}\int_{f_0}^{2f_0}|Z(f-f_0)|^2\langle F(f)F(-f+2f_0)\rangle_{Av}/Tdf. \quad (30)$$

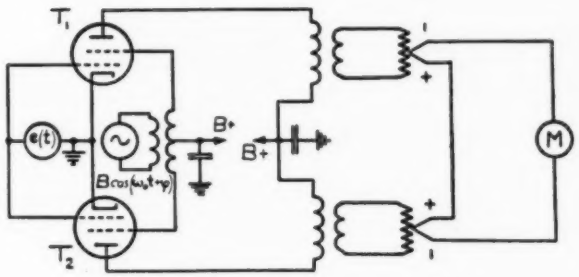


FIG. 1. Schematic of synchronous detector.

[7] W. Michels and N. Curtis, Rev. Sci. Instr. **12**, 444 (1941).

If

$$F_g(f)=\int_{-T/2}^{T/2} g(t)e^{-j\omega t}dt, \tag{31}$$

and if

$$G_g(f,f_0)=\langle F_g(f)F_g(-f+2f_0)\rangle_{Av}/T, \tag{32}$$

then it may be seen by direct substitution that for

$$g(t)=u(t)\cos\omega_k t, \quad (\omega_k\neq\omega_0)$$

$$\lim_{T\to\infty} G_g(f,f_0)=\lim_{T\to\infty} G_g(-f,-f_0)=0. \tag{33a}$$

If $g(t)=u(t)\cos\omega_0 t$, then

$$\lim_{T\to\infty} G_g(f,f_0)=\lim G_g(-f,-f_0)$$

$$=\lim_{T\to\infty}\langle|F_u(f-f_0)|^2\rangle_{Av}/4T. \tag{33b}$$

Hence the presence of noise modulated at the synchronous frequency yields a contribution to $\langle M^2\rangle_{Av}$ in addition to that furnished by the power spectrum of the voltage output from the receiver. Henceforth, we shall assume that $e(t)$ contains no interference of the form $u(t)\cos\omega_0 t$.

If $g(t)=A_k\cos\omega_k t$, $(\omega_k\neq\omega_0)$

$$\lim_{T\to\infty} G_g(f,f_0)=\lim_{T\to\infty} G_g(-f,-f_0)=0 \tag{34}$$

except for a discrete set of frequencies at which $\lim_{T\to\infty}G_g(f,f_0)$ is finite, so that $\int G_g(f,f_0)df$ vanishes over any region. For

$$g(t)=A_0\cos\omega_0 t,\ \lim_{T\to\infty} G_g(f,f_0)=\lim_{T\to\infty} G_g(-f,-f_0)$$

$$=\lim_{T\to\infty} A_0^2/4T\sin^2\frac{(\omega-\omega_0)T/2}{(\omega-\omega_0)^2/2}=\lim_{T\to\infty}|F_g(f)|^2/T. \tag{35}$$

Equations (33)–(35), together with the proviso of no noise modulated at the synchronous frequency, and the observation that $\lim_{T\to\infty}|F_g(f)|^2/T$ becomes infinite at $f=f_0$ only for $g=A_0\cos\omega_0 t$ permits Eq. (30) to be rewritten as follows:

$$\langle M^2\rangle_{Av}=D^2/4\int_0^{f_0-\epsilon}|Z(f-f_0)|^2P(f))df$$

$$+D^2/4\int_{f_0+\epsilon}^{2f_0}|Z(f-f_0)|^2P(f)df$$

$$+D^2/4\int_{f_0-\epsilon}^{f_0+\epsilon}|Z(f-f_0)|^2P(f)df, \tag{36}$$

where

$$P(f)=\lim_{T\to\infty} 2\langle|F(f)|^2\rangle_{Av}/T.$$

From Eqs. (27) and (35),

$$D^2/4\int_{f_0-\epsilon}^{f_0+\epsilon}|Z(f-f_0)|^2P(f)df$$

$$=\{D/2Z(0)A_0\}^2=\bar{M}^2. \tag{37}$$

Subtracting Eq. (37) from Eq. (36),

$$\langle(M-\bar{M})^2\rangle_{Av}=D^2/4\int_0^{f_0-\epsilon}|Z(f-f_0)|^2P(f)df$$

$$+D^2/4\int_{f_0+\epsilon}^{2f_0}|Z(f-f_0)|^2P(f)df. \tag{38}$$

Thus Eq. (37) states that the correct signal is obtained by integrating twice the power spectrum (multiplied by the gain characteristic, $[DZ(0)/2]^2$ over the neighborhood of the synchronous frequency. Equation (38) shows that the noise (or variance of the signal) is obtained by integrating the power spectrum (weighted by the corresponding gain characteristic) over the entire frequency range except in the neighborhood of $f=f_0$.

IV. CONCLUSION

An estimate of the minimum signal detectable from a receiver which extracts information from a periodic wave form in the presence of noise has been derived in terms of the first and second moments of an output meter reading. It was shown that these moments are obtainable from the power spectrum of the receiver output when (a) the meter is connected directly to the receiver, and (b) a synchronous detector with meter is used to extract the signal, subject to the condition that the receiver output wave form contain no noise modulated at the synchronous frequency.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 2 FEBRUARY, 1955

Patch Effect for the Thermionic Emission from Polycrystalline Tantalum*

W. B. LaBerge,† R. J. Munick,‡ J. A. Dezoteux, J. F. Whalen,§ and E. A. Coomes
University of Notre Dame, Notre Dame, Indiana

(Received February 1, 1954)

The slope modifications for the Schottky line implied by simple patch theory have been obtained at high fields for thermally etched tantalum filaments. For ac-aged filaments, only the usual low field patch modification was obtained. For dc-aged filaments an additional anomaly at higher fields was obtained, simultaneous with the appearance of the dc etch pattern on certain faces of the crystal grains. This high field anomaly, consisted of a double break in slope, and appeared at an average patch field about 50 times the patch field due to the distribution of crystal faces.

I. INTRODUCTION

PATCH theory[1] indicates that the experimental Schottky plot for the electron emission from a surface of nonuniform work function should experience a region of anomalous slope when the applied field becomes the same order of magnitude as the patch field. Becker,[2] and others[1] have observed this transition at low fields for the thermionic emission from thoriated tungsten; Apker[3] has obtained photoelectric data on semiconductor surfaces which exhibit the same type of patch effect. Schottky plots, taken with the precision required to separate the periodic deviations from the first-order Schottky effect, often contain more than one anomaly of slope. Figure 1 shows a high-field patch break for thermionic emission from tantalum. The periodic deviations are apparent in normal region III of these data.[4,5] The filament used here was a polished tantalum wire which had been vigorously outgassed by dc heating. Two aspects of this patch break are remarkable: (1) the anomaly occurs at a relatively high average patch field and (2) the width of the break is unusually narrow. The high value of patch field corresponds to small patches; the narrow patch break indicates an unusual patch distribution.[1]

An experiment was performed to isolate and study the high-field anomaly. Accurate Schottky plots were made for two groups of tantalum filaments. The first group, aged on ac, displayed only the usual low-field anomalous region. The second group, aged on dc, exhibited both the low-field anomaly and the high-field double break in slope exemplified by Fig. 1. Simultaneous with the high-field break, the dc etch pattern was observed on certain faces of the crystal grains.

II. EXPERIMENTAL METHOD

The experimental tube, measuring techniques, and experimental precision used here were described in a previous paper.[4] A variety of tantalum filaments were tested. These included 0.001-, 0.003-, and 0.005-inch unpolished and polished wires, with the same end results for all specimens. After the tubes were sealed off, Schottky plots were taken at various stages of thermal aging. Emission currents were measured at approximately 1500°K for a range of applied fields from 10^2 to 10^5 volts cm^{-1}. The temperature was held constant during any one run to within 0.02 percent. Coincident with the taking of Schottky data at various intervals of aging, changes in the structure of filament surfaces were followed by microscopic examination. The aging schedule consisted of approximate 6-hr intervals of heating in vacuum at 2500°K.

III. EXPERIMENTAL RESULTS

Figure 2 suggests the more prominent structure changes produced by thermal treatment. Two main effects are discernible: (a) On both ac and dc heating, crystal grains 1 to 10 wire diameters long and occupying the complete width grew along the filaments. Figure 2(a) shows this typical "bamboo" structure after 15 hours of heating on ac. While not fully formed after 5 hours, this structure nevertheless can be seen, and complete formation generally occurred within the

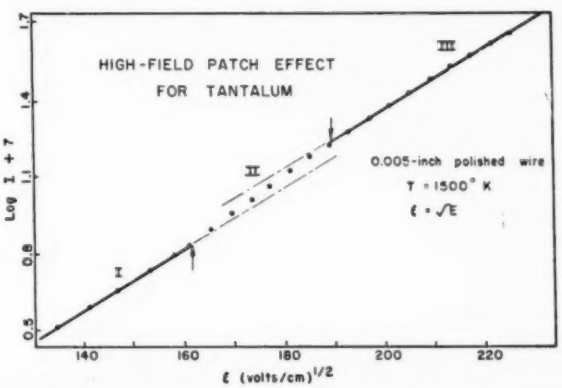


FIG. 1. Patch break at high fields in the Schottky plot for a polished tantalum filament. Vigorous outgassing had been carried out by dc heating.

* This research was sponsored by the U. S. Navy Bureau of Ships.

† Now at the U. S. Naval Ordnance Test Station, Inyokern, California.

‡ Now at National Argonne Laboratory, Chicago, Illinois.

§ Now at the Fournier Institute of Technology, Lemont, Illinois.

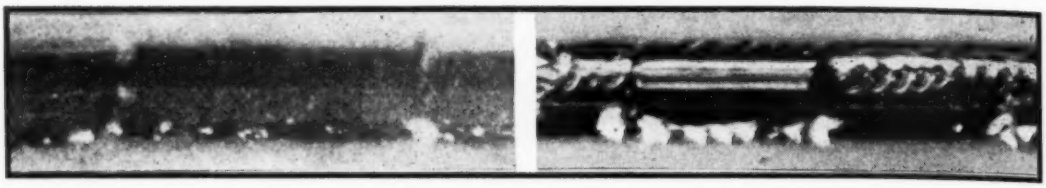
[1] See C. Herring and M. H. Nichols, Revs. Modern Phys. **21**, 185 (1949), Chapter 2, for review of theory and experimental data.

[2] J. A. Becker, Revs. Modern Phys. **7**, 95 (1935).

[3] See footnote *g*, reference 1.

[4] Munick, LaBerge, and Coomes, Phys. Rev. **80**, 887 (1950).

[5] LaBerge, Finn, and Coomes, Phys. Rev. **81**, 889 (1951).



(a) (b)

FIG. 2. Structure changes in 0.003 in. diameter tantalum filaments produced by heating at 2500°K. Filament (a) was heated on ac, for 15 hr and filament (b) on dc for 26 hr. Note the similarity of grain growth in both cases, and the etching produced by dc heating.

first 10-hr heating period. (b) On dc heating, the characteristic etching of various crystals occurred. While difficult to photograph on small cylindrical wires, the etch marks on various grain faces can be seen in Fig. 2(b), which is a photomicrograph taken after 26 hours of dc heating. The etch begins to appear clearly after about 10 hours dc heating, and the distinct hash-mark pattern reaches a stable state after about 25 hours. Visual observation reveals this to be identical with the pattern obtained for flat tantalum surfaces.[6] For the filament of Fig. 2(b) the etch marks are about $\frac{1}{5}$ wire diameter apart and form a configuration of steep hills separated by wide valleys. This is a good average picture for all dc heated specimens.

A representative family of Schottky plots taken on a filament at intervals during 26 hours of ac heating is shown in Fig. 3. For purposes of comparison, all Schottky lines have been displayed with a common slope. To obtain such a universal graph T, $\ln I$ *versus* ξ was plotted where I is the emission current and ξ the square root of the applied field.[7] The curves of Fig. 3 do not exhibit the double high-field break of Fig. 1, nor was etching of the filament surface observed during their run. The anomaly in this case is the usual patch effect found for thoriated tungsten, except that it occurs at lower fields. The break moves to still lower values of ξ as the grain size increases with aging time; for 25 hours of heating a value as low as 25 (volts $cm^{-1})^{\frac{1}{2}}$ has been obtained for some wires.

Figure 4 is representative of Schottky plots taken at intervals of dc-aged filaments. The low-field patch effects of Fig. 3 for ac heating are reproduced. In addition the high-field double break of Fig. 1 appears at $\xi_p \sim 150$ after 10 hours of aging. In contrast with the low-

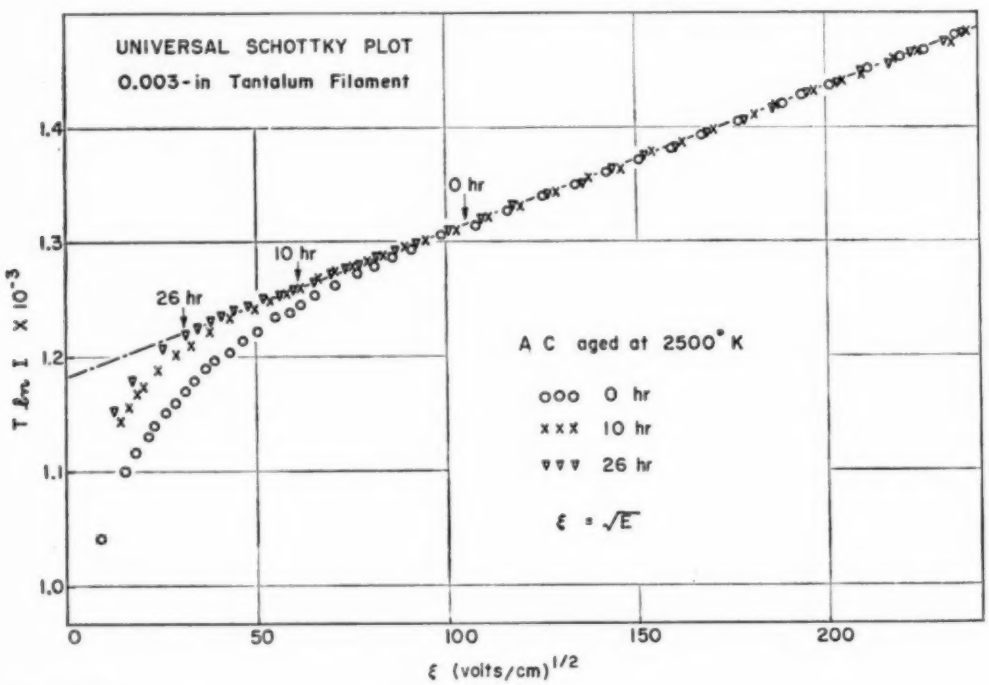


FIG. 3. Schottky plots for a tantalum filament aged on ac and taken at various steps in thermal treatment.

[6] B. Mrowca, J. Appl. Phys. 14, 648 (1943).

[7] The value of T was determined from the measured slope of the ordinary Schottky plot in the patch-free regions, where linearity resulted when the temperature was maintained constant. By this method the small uncertainties inherent in the experimental determination of the absolute temperatures of the various fila ments were absorbed. When this type of uncertainty has been eliminated from the Schottky plot, the patch anomalies are the only deviations that appear in prominence. See Figs. 3 and 4.

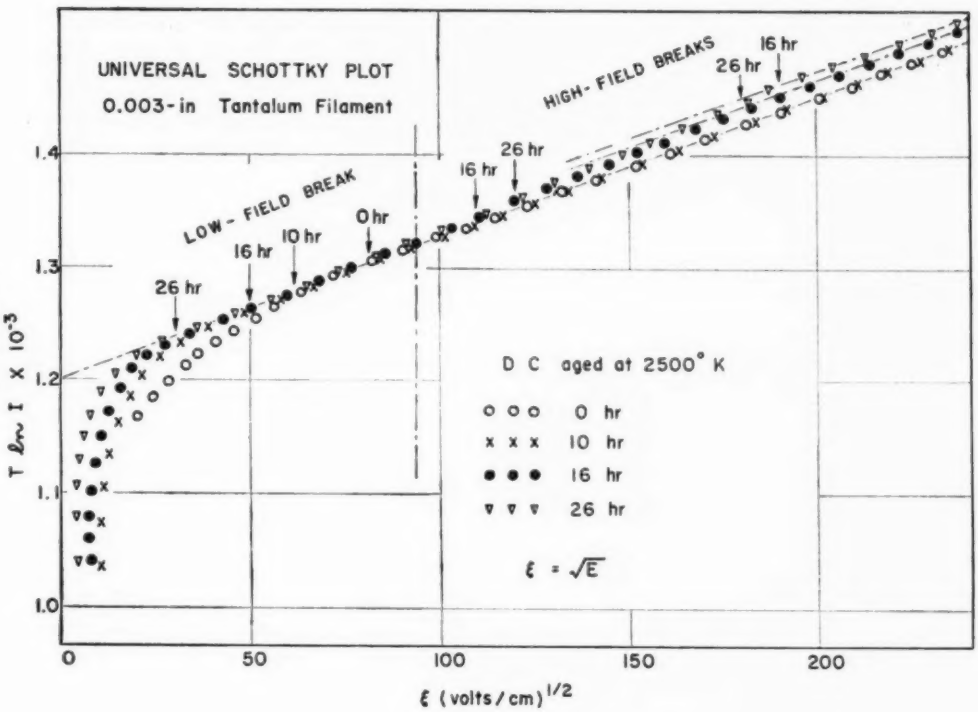


FIG. 4. Schottky plots for a tantalum filament aged on dc under conditions similar to those represented in Fig. 3. Note that the high-field patch modification of Fig. 1. first appears at the 16-hr interval.

field effect, ξ_p for this modification changes little with aging times. Values in the range $80<\xi_p<180$ have been observed on the various wires used in this experiment.

IV. DISCUSSION

While the precise values of ξ_p for both the low-field and high-field patch breaks vary somewhat from wire to wire, the two are always separable for tantalum filaments aged on dc. Table I gives a summary of average data for this experiment and a comparison with results published for thoriated tungsten. The comparison is based on the approximate theory of Herring and Nichols,[1] which states that the patch field $E_p \sim \delta\phi/\Delta y$, where $\delta\phi$ is the difference in work function of the patches, and Δy the patch diameter. Patch dimensions observed on photomicrographs have been recorded as (Δy) *observed*, while (Δy) *computed* for tantalum was obtained by comparing the data given in row *A* of the table with row *B*, and assuming $\delta\phi$ the same order of magnitude for both filament types. Order of magnitude agreement has been obtained for the observed and computed tantalum grain size, notwithstanding the rudimentariness of assumptions. Although the observed geometry for etch patterns differs radically from the configurations used for theory, the comparison between etch and grain patches for tantalum has been carried out in the table, lines *B* and *C*. For this comparison, (Δy) *computed* for the etches was based on (Δy) *observed* for the grains, while (Δy) *observed* for the etches was taken as the approximate width of the ridgetops observed on photomicrographs. One reason that the computed patches tend to smaller dimensions than observed etch marks may be that simple theory has not taken into account the sharpness of the ridges, which would tend to strengthen the patch fields over these areas.

Herring and Nichols[1] have remarked that the simple patch theory should apply to a superstructure of patches, and a configuration of this type should give the double break in Schottky slope shown in Fig. 1.

TABLE I. A comparison of patch effect data for thermally treated tantalum filaments and thoriated tungsten. Large crystal grains were formed in the tantalum by heating at 2500°K; certain crystal faces were etched by dc heating.

	Patch	Patch field (volts cm^{-1}) $\times 10^{-3}$	(Δy) *Observed* cm $\times 10^3$	(Δy) *Computed* cm $\times 10^3$
A	T_hW (grain)[a]	5	0.1–1	0.1–1
B	T_a (grain)	0.5	3–20	1–10[b]
C	T_a (etch)	25	0.2[c]	0.06–0.4[d]

[a] References 2 and 8.
[b] Computed from row A of Table, assuming $\delta\varphi$ constant.
[c] Width of etch patch taken as ½ etch spacing. See Fig. 2(b).
[d] Computed from row B of Table, assuming $\delta\varphi$ constant and using (Δy) *observed* for the tantalum grains.

[1] W. B. Nottingham, Phys. Rev. 49, 78 (1936).

Inertial Impaction of Aerosol Particles on Cylinders

J. B. Wong,* W. E. Ranz,† and H. F. Johnstone
University of Illinois, Urbana, Illinois

(Received February 19, 1954)

The inertial impaction of aerosol particles on cylinders was investigated to determine the relationship between the efficiency of impaction and the inertial parameter, the Reynolds number of flow, and the effect of interception. The impaction of sulfuric acid aerosols of nearly uniform size ranging from 0.6 to 1.4 microns diameter, on two platinum wires, 29 and 83 microns diameter, and two tungsten wires, 53 and 106 microns diameter, was measured at several velocities of flow past the wires. Reynolds numbers ranged from 13 to 330. For values of the square root of the inertial parameter $\Psi^{\frac{1}{2}}$ below 1.4, the experimental efficiencies of impaction agree with those predicted by Langmuir and Blodgett for potential flow and by Landahl and Herrmann for a Reynolds number of 10. The experimental values are higher than the calculated values for $\Psi^{\frac{1}{2}}$ greater than 1.4. The theoretical prediction that a critical value of $\Psi^{\frac{1}{2}}$ exists at approximately $\frac{1}{4}$, below which inertial impaction does not occur, was verified.

AN understanding of the mechanism by which aerosol particles are impacted on isolated cylinders is fundamental to the analysis of the process of filtration of dust and smoke. Practical analysis of the complex mechanics of particle motion may require consideration of inertia and interception, gravity settling, electrostatic attraction, Brownian diffusion, and thermal forces. However, all of these effects rarely occur simultaneously, and methods are available[1] for estimating the order of magnitude of each one. In the present work emphasis was placed on inertial impaction and the companion boundary effect of interception. Experimental data, which heretofore were obscure because of the wide spread in particle sizes of the test aerosol or unobtainable because of the difficulties of the experimental technique, are presented to support the contentions of theory.

MATHEMATICAL STATEMENT OF PROBLEM

Consider an aerosol stream moving in the x direction toward a circular cylinder placed perpendicular to the direction of flow. Consider also an aerosol particle moving with the stream. On approaching the collector the gas is deflected, but the relatively heavy particles tend to cross the stream lines and impact on the surface of the cylinder. At the same time, particle motion relative to the gas stream is slowed by fluid resistance, and the particle may fail to reach the surface before it is carried past the obstacle.

If all the forces except fluid resistance are inactive, application of the law of motion to the movement of the particle gives

$$-\frac{(\mathbf{u}-\mathbf{v})}{Z}=m\frac{d\mathbf{u}}{dt}. \tag{1}$$

The velocity of the particle relative to the gas is represented by $(\mathbf{u}-\mathbf{v})$. Z, the mobility, is defined in such a way that the forces opposing particle motion relative to the gas are given by $-(\mathbf{u}-\mathbf{v})/Z$. For lower relative velocities, and for particles in the range from 0.1 to 50 microns diameter, the Stokes-Cunningham resistance law applies and $Z=C/3\pi\mu D_p$. For higher velocities or larger particles, when the Reynolds number $[D_p|(\mathbf{u}-\mathbf{v})|\rho]/\mu$ is greater than about 0.1, Z becomes a complex function of the relative velocity and particle diameter. In the present work, only the range of the Stokes-Cunningham law is considered.

As a measure of the probability of a particle reaching the surface of the cylinder, the efficiency of inertial impaction η_I is defined as the ratio of the cross-sectional area of the original aerosol stream from which particle trajectories of particles of a given species and size intersect the surface to the projected area of the collector in the direction of flow. This definition implies that in any plane perpendicular to the direction of flow the number concentration of aerosol particles in the stream is uniform.

The equation of the particle trajectory in rectangular coordinates and dimensionless form is given from Eq. (1) by

$$\begin{aligned} 2\Psi\frac{d^2\bar{x}}{d\bar{t}^2}+\frac{d\bar{x}}{d\bar{t}}-\bar{v}_x=0,\\ 2\Psi\frac{d^2\bar{y}}{d\bar{t}^2}+\frac{d\bar{y}}{d\bar{t}}-\bar{v}_y=0, \end{aligned} \tag{2}$$

where $\Psi=mZv_0/D_c$, the inertial parameter, and $\bar{v}_x$ and $\bar{v}_y$ are the components of gas velocity which in the most general case considered here are functions of position $\bar{x}$ and $\bar{y}$ and Reynolds number N_{Re}. The boundary conditions for the particle at $\bar{t}=0$ and $\bar{x}=-\infty$ are $\bar{y}=\bar{y}_0$, $d\bar{x}/d\bar{t}=1$, and $d\bar{y}/d\bar{t}=0$. Thus, for a given particle and flow conditions, solution of Eq. (2) for the given boundary conditions results in a series of trajectories originating at values of $\bar{y}_0$ which intersect the surface of the cylinder or pass beyond it. One of these trajectories, originating at $\bar{y}_0{}^*$, is tangent to the cylinder, and it is this trajectory which determines the efficiency of impaction. In functional notation, the solution can be stated as follows: if

* Standard Oil Company (Indiana), Whiting, Indiana.
† Pennsylvania State University, State College, Pennsylvania.
[1] W. E. Ranz and J. B. Wong, Ind. Eng. Chem. 44, 1371 (1952).

$\bar{v}_x=\bar{v}_x(\bar{x},\bar{y},N_{\mathrm{Re}})$ and $\bar{v}_y=\bar{v}_y(\bar{x},\bar{y},N_{\mathrm{Re}})$, then $\eta_I=\bar{y}_0{}^*=\eta_I(\Psi,N_{\mathrm{Re}})$.

The inertial parameter Ψ has a physical significance in that it is the ratio of the stopping distance, i.e., the distance the particle will travel in still air when given an initial velocity of v_0, to the diameter of the cylinder. It is also the ratio of the constant force necessary to stop a particle initially traveling at a velocity v_0 in a distance $D_c/2$ to the fluid resistance on a particle traveling at a relative velocity of v_0. If the particle is spherical and $Z=C/3\pi\mu D_p$, $\Psi=C\rho_p v_0 D_p{}^2/18\mu D_c$. If C is nearly unity, $\Psi^{\frac{1}{2}}$ is directly proportional to particle diameter, and graphical representations of η_I as a function of $\Psi^{\frac{1}{2}}$ for various values of N_{Re} show directly the effect of particle diameter on inertial impaction.

TYPES OF FLOW

Elegant descriptions of the flow around a cylindrical obstacle, complete with boundary layer and separation lines, are available, but for the problem of impaction they are unnecessarily complicated and impractical. For the region in front of the cylinder where impaction takes place two simplified descriptions have been suggested, and they are probably adequate for a practical solution.

At high Reynolds numbers ($N_{\mathrm{Re}}\to\infty$) where the flow in front of the cylinder can be considered ideal, inviscid, irrotational, and incompressible,[2]

$$\bar{v}_x=1+\left[\frac{\bar{y}^2-\bar{x}^2}{(\bar{x}^2+\bar{y}^2)^2}\right],\qquad \bar{v}_y=\frac{-2\bar{x}\bar{y}}{(\bar{x}^2+\bar{y}^2)^2}. \tag{3}$$

At low Reynolds numbers a purely viscous solution, similar to Stokes' solution for spheres, is impossible, but Lamb's equation,[3] which takes partial account of the inertia terms in the differential equations of flow, is relatively simple and describes flow near the cylinder with good accuracy.

$$\bar{v}_x=\frac{1}{2.002-\ln N_{\mathrm{Re}}}[\ln(\bar{x}^2+\bar{y}^2)^{\frac{1}{2}}+(\bar{x}^2+\bar{y}^2-1)(\bar{y}^2-\bar{x}^2)/2(\bar{x}^2+\bar{y}^2)^2], \tag{4}$$

$$\bar{v}_y=-\frac{1}{2.002-\ln N_{\mathrm{Re}}}(\bar{x}^2+\bar{y}^2-1)\bar{x}\bar{y}/(\bar{x}^2+\bar{y}^2)^2.$$

Equation (4) is not applicable at large values of $\bar{x}$ and $\bar{y}$ since it does not reduce to $\bar{v}_x=1$ and $\bar{v}_y=0$ as $\bar{x}$ and $\bar{y}$ approach infinity.

Theoretically, Eq. (3) or (4) with a finite but relatively large starting position in the negative-x direction can be used in a numerical solution of Eq. (2) for limiting trajectories and impaction efficiencies. The calculations involved in such a procedure are extremely tedious.

MINIMUM INERTIAL PARAMETER BELOW WHICH PARTICLE TRAJECTORIES DO NOT REACH SURFACE

One of the interesting mathematical characteristics of Eq. (2) when Eq. (3) is used to describe the air velocity is that, for the region near the stagnation point, there exists a limiting solution which specifies a minimum Ψ of $\frac{1}{16}$ for a finite efficiency of impaction. For practical purposes this minimum Ψ can be interpreted as placing a lower limit on the size of aerosol particles whose trajectories reach the surface. Thus, it can be expected that very small particles tend to escape impaction all together. This condition is not entirely true, however, because the particles have a finite size and impact whenever their trajectories approach within a distance of $D_p/2$ of the collector surface. This effect is called interception.

MINIMUM EFFICIENCY OF IMPACTION DUE TO INTERCEPTION

For $\Psi=0$, e.g., a massless particle, there is a minimum efficiency of interception. For air flows described by Eq. (3), the minimum efficiency of impaction is determined by the $\bar{y}_0$ of streamlines which pass through $\bar{y}=1+R$ at $\bar{x}=0$ and is given by

$$\eta_c=(1+R)-\frac{1}{(1+R)}. \tag{5}$$

For flows described by Eq. (4), η_c is determined by the ratio of the flux of gas across the y axis between $\bar{y}=1$ and $\bar{y}=1+R$ to the total x-directed flux between $\bar{y}=0$ and $\bar{y}=1$ at $x=-\infty$. It is given by

$$\eta_c=[2(1+R)\ln(1+R)-R(2+R)/(1+R)]/2[2.002-\ln N_{\mathrm{Re}}]. \tag{6}$$

Equations (5) and (6) are plotted in Fig. 1.

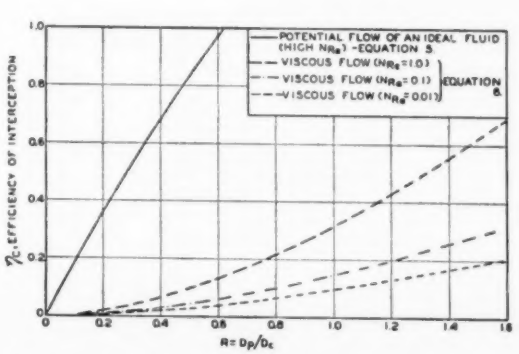


FIG. 1. Theoretical efficiency of interception on circular cylinders.

[2] H. Lamb, *Hydrodynamics* (Cambridge University Press, Cambridge, England, 1932), sixth edition, p. 77.
[3] Reference 2, 609–616.

NUMERICAL SOLUTIONS FOR EFFICIENCY OF IMPACTION DUE TO INERTIA

Several numerical solutions for impaction efficiency due to inertia have been obtained. Sell[4] showed that the efficiency of inertial impaction should be a unique function of the inertial parameter Ψ. He determined experimentally the velocity profile of water flowing past a circular cylinder 10 cm in diameter; and using this velocity distribution, he calculated stepwise the impaction efficiency curve shown in Fig. 2.

The work of Albrecht[5] was purely theoretical. He used Eq. (3) for his step-by-step calculations. His efficiency curve agrees with that of Sell for values of $\Psi^{\frac{1}{2}}$ greater than 0.4. For values of $\Psi^{\frac{1}{2}}$ smaller than 0.4, however, Albrecht's curve indicates a much greater slope and an intercept at $\Psi^{\frac{1}{2}}=0.3$. No explanation for the intercept was given.

Langmuir and Blodgett[6] employed the differential analyzer to establish curves both for the case in which Stokes' law is valid and for the case in which the complex drag coefficient of the particle must be considered. Their curve for the Stokes' law case is shown in Fig. 2. Theoretically, their results should be identical with those of Albrecht since the potential flow of an ideal fluid and Stokes' law were assumed in the derivations leading to both curves. The discrepancy is probably due to the fact that (a) Albrecht introduced the aerosol particle at the point $\bar{x}=-3$ and assumed that the particle velocity at that point is the same as that of the free stream, whereas Langmuir and Blodgett introduced the particle at $\bar{x}=-4$ and corrected the particle velocity by approximations, and (b) Albrecht had to resort to larger increments for his step-by-step calculations, whereas Langmuir and Blodgett used smaller increments with the differential analyzer. Langmuir and Blodgett recognized the existence of a minimum value of Ψ and give $\Psi^{\frac{1}{2}}=\frac{1}{4}$ below which inertial impaction does not take place.

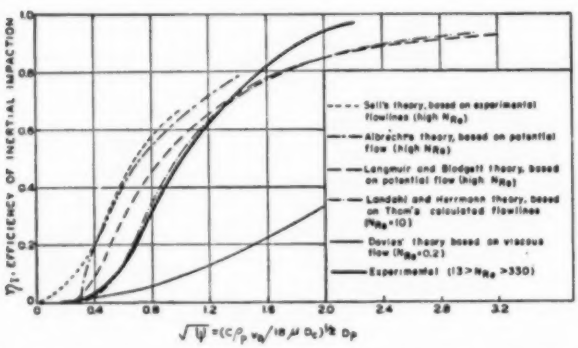


FIG. 2. Comparison of theoretical and experimental efficiencies of inertial impaction on circular cylinders.

Landahl and Herrmann[7] based their calculations on the flow field of Thom[8] for $N_{Re}=10$ and started their stepwise calculations with the particle at $\bar{x}=-6$, assuming the particle to be at the free stream velocity at that point. They fitted the resulting efficiencies with an empirical equation. The curve shown in Fig. 2 has been calculated from their equation. Although these authors did not recognize a critical value of Ψ, their empirical equation indicates that the efficiency of impaction is less than 0.001 for $\Psi^{\frac{1}{2}}=\frac{1}{4}$. They also present experimental data for the impaction of liquid droplets on a cylinder, but the data were not accurate enough to check the theory.

Davies,[9] using his own viscous flow equation[10] in which he had taken further account of the inertia terms in the differential equations of flow, calculated an impaction efficiency curve for $N_{Re}=0.2$. He predicted a much lower efficiency at low Reynolds numbers.

EFFICIENCY OF IMPACTION CAUSED BY INERTIA AND INTERCEPTION

The above information is sufficient to predict the form of the efficiency curve when both inertia and interception are taken into account. If η is plotted *vs* $\Psi^{\frac{1}{2}}$, there is a family of efficiency curves for each value of N_{Re} and range of R, or each value of R and range of N_{Re}. For high N_{Re}, the curves start at values of η given by Eq. (3) and are asymptotic to values of $\eta=1+R$ at $\Psi=\infty$. When N_{Re} is large and $R=0$, the curve begins at $\eta=0$, $\Psi^{\frac{1}{2}}=\frac{1}{4}$, and, because of this tendency to intercept the Ψ axis, has an S shape, characteristic of all inertial impaction curves.[1] At low Reynolds numbers, $N_{Re}<1$, the curves start at values of η approximated by Eq. (4) and rise only slowly with increasing $\Psi^{\frac{1}{2}}$. In general, for a given R and Ψ, η is much smaller at low values of the Reynolds number, a fact of considerable importance for the filtration of dust and smoke by fibrous filters.

EXPERIMENTAL DETERMINATION OF IMPACTION EFFICIENCIES

Figure 3 shows the experimental equipment used to establish the curve for the efficiency of inertial impaction of aerosol particles on circular cylinders. One-mil and 3-mil platinum wires and 2-mil and 4-mil tungsten wires were used as the cylinders. The diameter of each wire was measured accurately under the microscope. Table I shows the arithmetic averages and maximum deviations from the average.

The assembly consisted of a brass drum from which the wire was unwound, a horizontal Venturi-shape nozzle made of Lucite through which the aerosol stream passed, and an 8-mm Pyrex-glass wash tube into which the wire which had been exposed could be lowered and

[4] W. Sell, Forsch. Gebiete Ingenieurw. **2**, *Forschungsheft* 347 (1931).
[5] F. Albrecht, Physik. Z. **32**, 48 (1931).
[6] I. Langmuir and K. B. Blodgett, "Mathematical investigation of water droplet trajectories," General Electric Research Laboratory, Schenectady, New York, Rept. No. RL225, 1944–1945.
[7] H. D. Landahl and R. G. Herrmann, J. Colloid Sci. **4**, 103 (1949).
[8] A. Thom, Proc. Roy. Soc. (London) **A141**, 651 (1933).
[9] C. N. Davies, Proc. Inst. Mech. Engrs. (London) **B1**, 185 (1952).
[10] C. N. Davies, Proc. Phys. Soc. (London) **B63**, 288 (1950).

washed with conductivity water. The drum was attached to the center of a circular brass plate. The winding and unwinding of the wire was accomplished by manually rotating the plate in the appropriate direction. The drum was enclosed in a Lucite casing to prevent the wire from being contaminated.

The Lucite nozzle was installed with its circular cross section in a vertical plane directly below the brass drum. The throat of the nozzle was 0.357 cm in diameter and approximately 0.6 cm long. The cone-shape convergent and divergent sections were both tapered at an angle of 60°. Small holes were drilled through the upper and lower walls of the nozzle throat to allow the wire to pass down into the wash tube. A brass weight, ranging from 1 to 4 g, depending on the wire diameter, was used to plumb the wire.

The homogeneous aerosol generator was similar to that used by Sinclair and LaMer.[11] The air for the generator was supplied at 90 lb per sq in. pressure, saturated with water vapor, and filtered through a porous plate. It was then passed through a pressure regulator and an Army Chemical Corps aerosol filter canister (Filter, Special, Particulate, E12R3) to two

TABLE I. Wire diameters.

Wire	Average diameter microns	Maximum deviation from average, percent
Platinum, 1 mil	29.0	8.6
Platinum, 3 mil	82.6	3.6
Tungsten, 2 mil	53.1	6.1
Tungsten, 4 mil	105.7	0.87

rotameters. One rotameter delivered air to the salt nuclei generator while the other delivered air to the bubbler of the aerosol generator.

Condensation nuclei were generated from a mass of sodium chloride held on a spiral Nichrome ribbon contained in a glass flask, which was heated by a current ranging from 2.5 to 3.2 amp. The current was controlled manually by a Variac in conjunction with an ammeter and a voltage regulator.

The aerosol generator was enclosed in an insulated box. The two glass flasks, one containing the aerosol material and the other serving as a reheater, were placed in metallic tanks filled with liquid paraffin heavy oil. The oil baths were heated electrically by means of Calrod heaters installed directly below them. The temperatures in the baths were regulated by mercury thermoregulators.

Concentrated sulfuric acid, C.P., 95.5 percent, was used as the aerosol material. The generator produced aerosol particles of nearly uniform diameters in the range from 0.6 to 1.4 microns at a rate of 1 to 5 mg per minute.

The aerosol was pulled through the equipment by means of a rotary vacuum pump. It was mixed with

[11] D. Sinclair and V. K. LaMer, Chem. Revs. 44, 245 (1949).

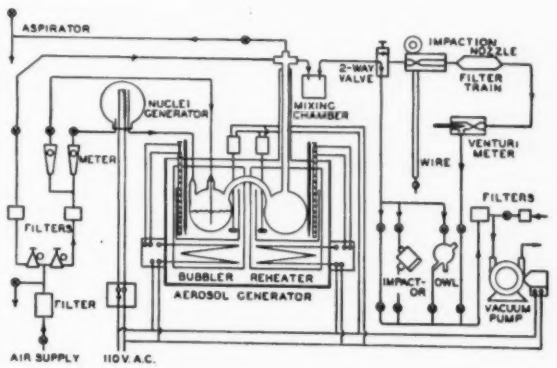


FIG. 3. Experimental equipment for determining impaction efficiency of circular cylinders.

filtered air to obtain the desired rates of flow through the nozzle. From the mixing chamber, it passed through a two-way quick-opening valve. When the valve was in one position, the aerosol flowed through one of the following paths: (a) through a line leading directly to the filter before the vacuum pump; (b) through a high velocity impactor for the determination of the acid concentration of the particles; or (c) through the "Owl" for the measurement of particle diameters.

After the Lucite nozzle, the aerosol stream passed through a filter train and a Lucite Venturi meter before reaching the vacuum pump. The filter train consisted of three filters, each packed with one inch of No. 800 Pyrex Brand wool filtering fiber. These could be removed and weighed separately. Complete removal of the particles was assumed if nearly all the weight increase appeared in the first two filters. The flow rate through the nozzle was determined by two calibrated Lucite Venturi meters with different ranges of flow rates. A mercury thermometer was installed at the downstream end of the Venturi meter. The temperature indicated by this thermometer was always nearly the same as the room temperature and was taken as the temperature at the nozzle throat.

The acid concentration of the aerosol particles was determined by collecting a weighable quantity of particles in the cup of a high-velocity impactor[12] and analyzing the content in the cup by titration with 0.1*N* sodium hydroxide solution.

The diameter of the aerosol particles was measured with a calibrated "Owl" (No. G-2)[13] obtained from the U. S. Army Chemical Corps. When additional air of nearly the same moisture content was mixed with the aerosol rising from the generator to give the necessary velocity through the nozzle throat, the particle diameter remained substantially the same as that before the air was added.

A precision conductivity bridge built according to the

[12] W. E. Ranz and J. B. Wong, Arch. Ind. Hyg. Occupational Med. 5, 464 (1952).

[13] D. Sinclair, "Measurement of Particle Size and Size Distribution," in *Handbook on Aerosols* (U. S. Atomic Energy Commission, Washington, D. C., 1950), p. 97.

design of Luder[14] was used to analyze the quantity of acid particles impacted on the wire. A dip conductivity cell with platinized platinum electrodes with a cell constant of approximately 0.1 was employed. The bridge was capable of indicating down to 0.02 percent of the cell resistance. A Simpson meter was used for balancing the bridge. The dip cell was calibrated at 30°C using 10^{-6} to $10^{-2}N$ sulfuric acid solutions made by diluting $10^{-1}N$ standard sulfuric acid solutions.

PROCEDURE

The aerosol generator was operated for about four hours before experimental runs were started. This allowed the operating conditions to be adjusted and steady state to be established. During the later part of this period, the vacuum pump was started and samples of the aerosol were taken in the "Owl." Particle size measurements were made every five to ten minutes until the same size was indicated in successive readings.

As soon as the particle size had been established, the two-way valve was changed to allow the aerosol stream to pass through the nozzle for measuring the impaction of particles on the vertical wire hung across the nozzle throat. The valve regulating the input of additional air and those regulating the input to the vacuum pump were adjusted until the desired flow rate through the nozzle was obtained. The circular plate attached to the brass drum was rotated manually to lower the wire at a predetermined rate so that enough particles would be collected on the wire for accurate analysis but not enough to cause re-entrainment of particles from the surface of the wire. The time for approximately 4 feet of the wire to pass the nozzle throat ranged from 3 to 19 minutes depending on the estimated rate of aerosol generation and the efficiency of impaction for that test run. The duration of most runs was 4 to 6 minutes.

At the completion of a run, the aerosol was allowed to pass the "Owl" so that the particle size could be rechecked. In most runs, the particle size did not change appreciably during the run. The quantity of aerosol which escaped impaction was determined by weighing the filters in the filter train before and after the run.

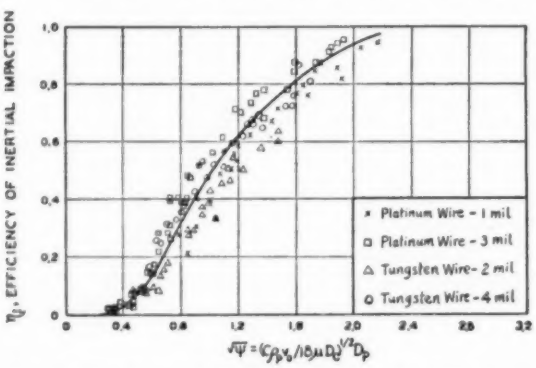


FIG. 4. Experimental efficiencies of inertial impaction on circular cylinders.

[14] W. F. Luder, J. Am. Chem. Soc. **62**, 89 (1940).

The acid collected on the wire was washed with a known quantity of conductivity water (45 cc for most runs). The solution was withdrawn through the bottom of the tube, and the washing procedure repeated twice with the same solution. The solution was divided into three tubes and placed in the thermostat at 30°C. The dip conductivity cell was thoroughly washed in one of the tubes, and the concentrations of the solutions in the other two tubes were measured. For most runs, the concentrations of the two solutions checked with each other within the accuracy of the calibration curve for the cell.

For each series of runs made under one set of conditions, the acid concentration of the aerosol particles was determined by collecting the particles in an impactor cup, using a velocity sufficient to give 100 percent removal efficiency.[12] The density of the particles was obtained from tables.

EXPERIMENTAL IMPACTION EFFICIENCY CURVE

Figure 4 shows the experimental impaction efficiencies of sulfuric acid aerosol particles on the four sizes of wires. The ranges of variables represented are: diameter of aerosol particles, 0.56 to 1.40 microns; superficial velocity of aerosol stream passing wires, 399 to 5090 cm/sec; diameter of wires, 29.0, 53.1, 82.6, and 105.7 microns; and Reynolds number based on the measured wire diameters, 13.0 to 330. The acid concentration of the particles averaged about 58.5 percent sulfuric acid with a maximum deviation of less than 6 percent. The density of the aerosol particles was substantially constant at 1.48 g/cc.

In the range of particle diameter and aerosol stream velocity employed in the experiments, collection due to Brownian diffusion was negligible. Collection by electrostatic forces was improbable since both the aerosol particles and the wires were uncharged. Gravity settling should also be unimportant with the wires in the vertical position. The values of the interception parameter R ranged from 0.022 to 0.038 for the 29.0 micron wire, from 0.013 to 0.023 for the 53.1 micron wire, from 0.007 to 0.017 for the 82.6 micron wire, and from 0.005 to 0.012 for the 105.7 micron wire. The effect of interception was considered to be negligible for the two large wires. For the 29.0 and 53.1 micron wires, impaction efficiencies recorded were, respectively, 0.21 and 0.074 or higher, so that small effects of interception would not appreciably affect the shape and position of the impaction efficiency curve. The curve drawn through the points in Fig. 4, therefore, represents experimental efficiencies of inertial impaction for a relatively high range of Reynolds numbers.

The experimental curve indicates a critical value of $\Psi^{\frac{1}{2}}$ of approximately 0.25, below which inertial impaction does not occur. At higher values of the inertial parameter the curve appears to be asymptotic to the value of $\eta_I = 1$. The impaction efficiencies were reproducible in

terms of $\Psi^{\frac{1}{2}}$, in view of the tenfold variation in velocity, the threefold variation in particle size, and the insensitivity of the "Owl" for detecting small variations in the particle size. The accuracy and reliability of the results depend largely on the homogeneity of the particle size and the accuracy of the particle size measurements.

EFFECT OF REYNOLDS NUMBER

For the two platinum wires, for which the values of the Reynolds number, N_{Re}, ranged from 13.7 to 32.8 for the 1-mil wire and up to 248 for the 3-mil wire, the impaction efficiency in all cases is higher for higher values of N_{Re} with the same value of $\Psi^{\frac{1}{2}}$. The same observation can be made on the two tungsten wires, for which N_{Re} ranged from 13.0 to 58.5 for the 2-mil wire and up to 330 for the 4-mil wire. This agrees with the conclusion drawn previously.

Comparison of the data for the 1-mil platinum wire with those for the 2-mil tungsten wire, and the data for the 3-mil platinum wire with those for the 4-mil tungsten wire, however, show an opposite effect of the Reynolds number; e.g., the impaction efficiencies on the 3-mil wire at lower values of N_{Re} are generally higher than those on the 4-mil wire for corresponding values of $\Psi^{\frac{1}{2}}$. The explanation of this is not apparent. When these wires were examined under the microscope, it was noted that the surface of the platinum wire was much smoother than that of the tungsten wire. Possibly the aerosol particles adhered to the surface of the platinum better than to that of the tungsten wire.

COMPARISONS WITH THEORETICAL CURVES

Figure 2 is a plot of the experimental inertial impaction efficiency curve together with the various theoretical curves proposed. The experimental curve agrees closely with that calculated by Landahl and Herrmann based on $N_{Re}=10$, up to $\Psi^{\frac{1}{2}}=1.4$. For values of $\Psi^{\frac{1}{2}}$ between 0.4 and 1.2, the experimental data indicate impaction efficiencies somewhat smaller than those shown by the curve of Langmuir and Blodgett and considerably smaller than the values of Sell and of Albrecht. This is to be expected since the curves of Langmuir and Blodgett and of Albrecht were based on the potential flow of an ideal fluid, and that of Sell was based on a flow pattern at large values of N_{Re} on a 10-cm cylinder. In the experimental flow conditions, viscous effects enter to some extent and decrease the impaction efficiencies below their values for ideal flow.

For values of $\Psi^{\frac{1}{2}}$ greater than about 1.4, the experimental efficiencies are higher than those according to the curves of Langmuir and Blodgett and of Landahl and Herrmann. The reason for this discrepancy is not entirely clear. One point to be noted, however, is that, for higher efficiencies of inertial impaction, i.e., efficiencies approaching unity, the particle trajectories are nearly parallel to the direction of flow and the particle must cut across the streamlines far upstream of the cylinder where the streamlines begin to spread. In stepwise calculations of particle trajectories, it is not practical to start the calculation more than a few diameters of the collector upstream. This method was followed by both Langmuir and Blodgett, and Landahl and Herrmann. It is possible that errors introduced by this could cause the calculated efficiencies in the high efficiency range to be lower than the correct values.

No comparison can be made with Davies' theoretical curve since it was based on viscous flow at $N_{Re}=0.2$, far below the range attainable with the method used in the experiment.

The critical value of $\Psi^{\frac{1}{2}}$ at approximately 0.25 shown by the experimental impaction efficiencies agrees with the values of 0.3 and 0.25 explicitly stated by Albrecht, and Langmuir and Blodgett, respectively, and substantiates an important theoretical prediction.

ACKNOWLEDGMENT

This work is a part of an investigation on the fundamental properties of aerosols being conducted in the Engineering Experiment Station of the University of Illinois under Contract Number AT(30-3)-28 with the U. S. Atomic Energy Commission.

NOMENCLATURE

C = Empirical correction factor for resistance of gas to the movement of submicron particles, dimensionless. $C=1+(2\lambda/D_p)[1.23+0.41\exp(-0.44D_p/\lambda)]$ for $0.1<2\lambda/D_p<134$.
D_c = diameter of cylinder, cm.
D_p = diameter of aerosol particle, cm.
m = mass of aerosol particle, g.
$N_{Re}=D_c v_0\rho/\mu$ = Reynolds number of cylinder.
R = D_p/D_c = interception parameter, dimensionless.
t = time, sec.
$\bar{t}$ = $2v_0t/D_c$, dimensionless time.
$\mathbf{u}$ = velocity of aerosol particle, cm/sec.
$\mathbf{v}$ = velocity of fluid, cm/sec.
$\bar{v}_x$ = v_x/v_0, $\bar{v}_y=v_y/v_0$, dimensionless fluid velocity components.
v_0 = average, unidirectional velocity of the fluid far upstream of the cylinder, cm/sec.
x, y = position coordinates, cm.
$\bar{x}$ = $2x/D_c$, $\bar{y}=2y/D_c$, dimensionless position coordinates.
$\bar{y}_0$ = original $\bar{y}$ position of particle at $\bar{t}=0$, dimensionless.
$\bar{y}_0^*$ = $\bar{y}_0$ of particle whose trajectory is tangent to cylinder, dimensionless.
Z = mobility of particles, cm/dyne sec.
η = efficiency of impaction, dimensionless.
η_C = efficiency of interception, dimensionless.
η_I = efficiency of inertial impaction, dimensionless.
ρ = density of fluid, g/cc.
ρ_p = density of particle, g/cc.
Ψ = mZv_0/D_c = inertial parameter, dimensionless.
λ = mean free path of gas molecule, cm.
μ = viscosity of fluid, poise.

Letters to the Editor

Wall Effects on Microwave Measurements of Ferrite Spheres

E. G. SPENCER AND R. C. LECRAW
Diamond Ordnance Fuze Laboratories, Washington, D. C.
(Received August 11, 1954)

IN measuring the microwave properties of ferrites, it is convenient to use small spherical samples in high Q resonant cavities. The internal rf electric and magnetic fields can be calculated and thus precise measurements of electric and magnetic susceptibilities can be obtained. One assumption needed to calculate the internal fields is that the applied rf field lines are parallel over the sample. The magnitude of the field inside the spherical sample is then a constant;[1] however outside the sphere the field lines are not parallel for an appreciable distance, and if the sphere is placed on the cavity wall for magnetic measurements any rf magnetic field component normal to the wall generates extraneous eddy currents in the wall. These give rise to an additional unevaluated magnetic field which is nonuniform over the sample. The total rf internal magnetic field becomes quite complicated and causes any conclusions based on the assumption of uniform magnetic fields to be in error.

A similar discussion holds for electric field measurements. Placing the sphere against a conducting wall results in rf electric field components parallel to the wall causing a reaction which complicates the rf field inside the sphere. We have found experimentally that the usual type frequency shift measurements made on a ferrite sphere placed on the wall of a resonant cavity can result in a frequency shift 25 percent higher than if the sphere is placed away from all walls.

For the magnetic measurements, Fig. 1 shows a typical frequency shift for a rectangular cavity of $Q=2500$ as a function of H for a 1.25 mm diameter sphere of a magnesium-manganese ferrite,[2] the sample being properly placed away from all walls and in a position of the maximum rf magnetic field. Figure 2 shows the variation in the difference between maximum ΔF positive and maximum ΔF negative as the sample is lowered equidistant from the cavity walls in a long rectangular cavity. The disturbing effect of the end wall is clearly indicated. Thus for accuracy of electric and magnetic microwave measurements of any kind the spherical sample should be away from the wall.

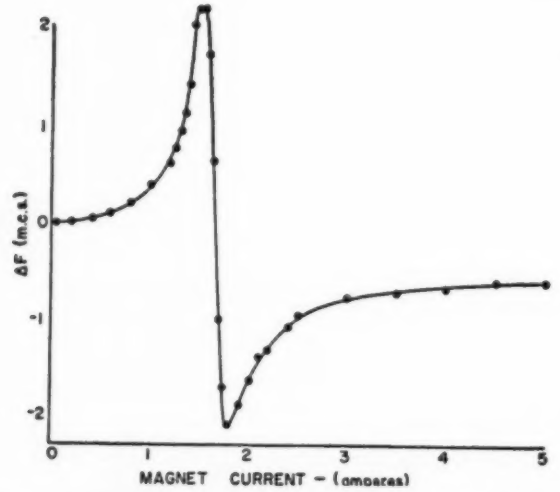


FIG. 1. Frequency shift of the resonance of the cavity containing a 1.25 mm sphere of a magnesium-manganese ferrite, as a function of the steady magnetic field.

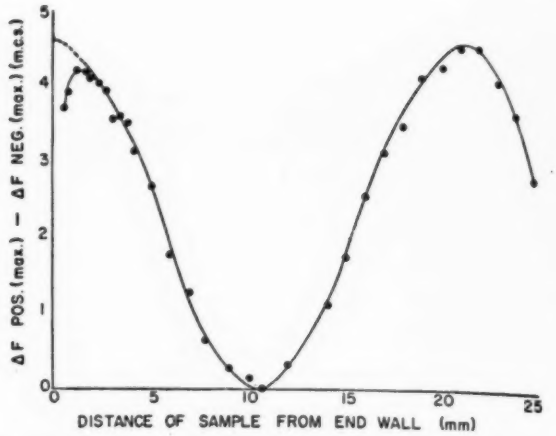


FIG. 2. Difference between ΔF positive (maximum) and ΔF negative (maximum) (as shown in Fig. 1) as a function of distance of sample from the end wall. The dashed line is a mirror image of the second half of the curve and indicates the region in which wall effects are appreciable.

Another wall effect to be considered is caused by the induced rf field $h_y{}^i$ normal to the applied rf magnetic field $h_x{}^0$, where the superscript i refers to internal fields and $h_y{}^i$ for a spherical sample is given by

$$h_y{}^i=\frac{-3iK}{(\mu+2)^2-K^2}h_x{}^0.$$

Thus, $h_y{}^i$ which comes in through the off-diagonal components of the permeability tensor couples in an rf magnetic field which in a circular or square cavity would be a mode normal to the applied microwave mode. In a rectangular wave guide cavity, the wavelength of this mode would be greater than cutoff. There would be an entirely different effect on the magnetic poles on the surface of the sphere placed in the two cavities, and the reaction in the rectangular one would cause the internal fields to be given some other form than the above equation.

The magnitude of this effect would depend on the ratio of the volume of the sample to the volume of the cavity. This ratio is required to be small if the perturbation techniques in calculations are to be used. No experimental check on the magnitude of this effect has been carried out; however, with an additional wall effect for the rectangular cavity, it would appear that better accuracy would be obtained for cylindrical or square cavities than could be obtained in rectangular cavities assuming the same sample to cavity volume ratio.

The authors wish to thank F. Reggia for his assistance in the measurements, and R. D. Hatcher for his interest in this problem and for his many helpful discussions.

[1] H. G. B. Casimir, Philips Research Reports 2, 34 (1947).
[2] The magnesium-manganese ferrite is manufactured by General Ceramics and Steatite Corp. Keasby, New Jersey. It is called 1331.

The Transition Structure of Nickel

G. I. FINCH, K. P. SINHA, AND A. GOSWAMI
National Chemical Laboratory of India, Poona 8, India
(Received July 16, 1954)

NICKEL exhibits anomalies in its physical properties such as specific heat, temperature coefficients of expansion and electrical resistance, and rigidity etc., in the range of temperature between 300°C and 370°C.[1] Its Curie point lies at 358°C. These anomalies cannot be fully accounted for by the magnetic transformation associated with the Curie point.[2] In certain ferromagnetic alloys changes in crystal structures also occur near the

FIG. 1. Reflection pattern of nickel film on glass deposited at 300°C showing mostly rings of b.c.c. nickel.

Curie point,[3] but no such phase transformation seems to have been observed for nickel within this region. Normally its stable structure below and above this temperature range is face-centered-cubic but the existence of hexagonal (close-packed) structure has also been reported in a few cases.[4–6] No body-centered-cubic form has so far been observed in evaporated films. Recently, however, we have found by electron diffraction that under appropriate conditions a body-centered-cubic nickel can be formed in thin films which on heating to beyond 400°C goes over to the normal face-centered-cubic structure.

In our experiments nickel was thermally deposited from a nickel-coated hot tungsten wire in a high vacuum at various temperatures onto heated substrates such as the cleavage faces of rock salt and mica, or on glass surfaces, situated about five centimeters away from the hot filament. The temperature of the substrate was measured with a calibrated thermocouple. After evaporation for a few seconds, the specimens were cooled rapidly *in vacuo* to room temperature and transferred to the electron-diffraction camera. Blank experiments with an uncoated tungsten wire under otherwise similar conditions did not result in the formation of any tungsten film detectable by electron diffraction, suggesting that the rate of evaporation of tungsten compared with that of the nickel coating was small or even negligible.

The translucent nickel films obtained by deposition on to any of the above substrates heated to between 300°C and 340°C revealed a body-centered-cubic structure (Fig. 1) of which $a_0=2.78$ A by calibration against the $11\bar{2}0$ ring of graphite. This structure was found to be metastable in that after heating for one hour at 400°C in a high vacuum, followed by rapid cooling, it was found to have reverted to the normal f.c.c. structure of nickel (Fig. 2) with $a_0=3.52$ A.

None of the electron diffraction patterns obtained in these experiments showed any trace of a preferred crystal orientation. From this absence of any sign of epitaxy it may be concluded that the film growth was unaffected by the nature of the substrate. It would seem therefore that the formation of the b.c.c. phase was primarily dependent on the substrate temperature, the structure being b.c.c. between about 300°C and 340°C, but f.c.c. outside this range. The rate and duration of evaporation affected the ease with which the b.c.c. nickel structure could be observed, but this would seem to be a secondary effect for which film thickness and crystal size were responsible, the metastable b.c.c. structure being most easily frozen in thin films and small crystals.

The entropy must increase on going from the f.c.c. to b.c.c. structure owing to the greater freedom of oscillation in the b.c.c. structure, which has a lower coordination number than that of the f.c.c. structure.

At present it cannot be ruled out that traces of foreign atoms, such as tungsten, may play an important role in stimulating the formation of the metastable b.c.c. phase. Micro-crystals of tungsten (b.c.c.) might well act as potential centers of crystallization of the new phase and enhance the nucleation rate. In other words, these microcrystals may act as embryos in the matrix of the b.c.c. phase.[7]

FIG. 2. Reflection pattern of the same specimen after heating for one hour at 400°C in high vacuum showing transformation to f.c.c. nickel.

Another point of interest is that in the case of nickel the specific heat *vs* temperature curve, though exhibiting a sharp hump in the anomalous region, is continuous thus showing that there is no λ anomaly. The rise in specific heat can be correlated both with the structural as well as with the magnetic transformations. A detailed discussion will be published later.

We are thankful to Dr. A. B. Biswas for helpful discussions.

[1] J. W. Mellor, *A Comprehensive Treatise on Inorganic and Theoretical Chemistry* (Longmans Green and Company, London, 1936), Vol. XV.
[2] E. C. Stoner, Repts. Prog. Phys. 11, 43 (1948).
[3] R. M. Bozorth, *Ferromagnetism* (D. Van Nostrand Company, Inc., New York, 1951), p. 715.
[4] G. P. Thomson, Nature 123, 912 (1929).
[5] Finch, Wilman, and Yang, Disc. Faraday Soc. 43A, 144 (1947).
[6] G. Le Clere and A. Michel, Compt. rend. 208, 1583 (1939).
[7] R. Smoluchowski, *Phase Transformation in Solids* (John Wiley & Sons, Inc., New York, 1951), p. 149.

The Effect of Superposing a Small Alternating Excitation on the Steady Excitation of a Luminescent Material

KENNETH F. STRIPP* AND RICHARD H. BUBE
Radio Corporation of America, RCA Research Laboratories, Princeton, New Jersey

(Received October 23, 1954)

THE somewhat novel experiment involving the superposition of a small alternating excitation (in either beam current or beam voltage for cathodo-excitation) on the steady excitation of a luminescent material, leads to a number of useful relations between characteristic parameters of luminescence excitation and decay.

Assuming a layer of luminescent material of a thickness x_0, the excitation density in the phosphor may be calculated if it is known how the energy of the beam is dissipated in penetrating the phosphor layer. The penetration function of Dowling and Sewell[1] is assumed in this paper as

$$E=iV\exp(-x/KV), \qquad (1)$$

where E is the beam power per cm² incident at depth x, i is the incident current density at $x=0$, x is the penetration parameter, V is the accelerating voltage, and KV is equivalent to the range of the exciting beam. If Q is the efficiency of energy transfer, the excitation intensity per unit volume, $L(x)$, may be calculated such that

$$L(x)=-QdE/dx=i(Q/K)\exp(-x/KV). \tag{2}$$

Let it be assumed that the excitation, $L(x)$, is given by the sum of a steady excitation, $L^*(x)$, and a small alternating excitation. First the case of an alternating perturbation on the beam current is considered.

$$L(x,t)=L^*(x)(1+\epsilon\sin\omega t). \tag{3}$$

The perturbation in excitation current will lead to a perturbation in the concentration of excited centers:

$$N(x,t)=N^*(x)[1+\epsilon z(x,t)] \tag{4}$$

where $N^*(x)$ is the concentration of excited centers for steady excitation alone. If Eqs. (3) and (4) are substituted in the general equation for a first-order luminescent material such that

$$\frac{dN(x,t)}{dt}=L(x,t)[N_0-N(x,t)]-kN(x,t). \tag{5}$$

Where N_0 is the total concentration of effective centers and k is the decay constant. It is now possible to derive a solution for the luminescence emission intensity

$$I=I^*[1+\epsilon A\sin(\omega t-\phi)] \tag{6}$$

where I^* is the emission intensity for steady excitation, A is related to the amplitude of the alternating component of the emission, and ϕ is the phase shift between the alternating emission and the alternating excitation. General expressions for A and ϕ may be derived, but it is more useful to consider the following approximations: (1) For a low steady excitation intensity, i.e., $L(x)\ll k$, it is found that:

$$A_0=k/(k^2+\omega^2)^{\frac{1}{2}} \quad \text{and} \quad \tan\phi_0=\omega/k \tag{7}$$

where the zero subscript indicates the low intensity approximation. It may also be shown that:

$$A=A_0(1-[2k/(k^2+\omega^2)]L\cdots) \tag{8}$$

which indicates that A approaches A_0 at high frequencies. Since, from Eqs. (7) and (8) it is possible to express.

$$\frac{A_0-A}{2A_0{}^3}=L/k=R, \tag{9}$$

a plot of R as a function of L will have a slope of $1/k$.

(2) For an intermediate range of steady excitation intensities, it is found that the phase shift may be written as

$$\tan\phi=\omega[k+(iQ/2K)\exp(-d/KV)]^{-1} \tag{10}$$

where d is the thickness of the "dead layer" at the surface of the phosphor particles. The thickness of the "dead layer" enters in this approximation as the result of the integration over x from d to x_0 to get the total emission. Since the thickness of the "dead layer" enters through $L(x)$, according to Eq. (2), it did not appear in the first approximation where $L(x)$ was assumed negligible with respect to k.

If $\tan\phi$ is plotted as a function of frequency, the slope determined from the low-frequency end, may be obtained as a function of current, i. If the inverse slope is plotted as a function of i, the result will be a straight line with intercept k and slope $(Q/2K)\exp(-d/KV)$. If this experiment were performed for different values of voltage, it would also be possible to check the Dowling and Sewell approximation of the penetration function, as well as to determine values for d/K and $Q/2K$.

(3) For a high steady excitation intensity, the phase shift may be expressed as

$$\tan\phi=\omega[2k(1-k/L(d)]^{-1} \tag{11}$$

where $L(d)$ is a short notation for $(iQ/K)\exp(-d/KV)$. Thus the inverse slope of a plot of $\tan\phi$ *vs* frequency, determined from the low-frequency end, should approach the value of $2k$ for large values of excitation intensity.

(4) For high alternating frequency, i.e., $\omega\gg(k+L)$, the phase shift may be expressed such that

$$\tan\phi=\omega\ln[1+L(d)/k]L^{-1}(d). \tag{12}$$

(5) If, instead of limiting the range of frequency and excitation intensity, the thickness of the sample is limited, so that it is possible to consider essentially uniform excitation throughout the thickness of the sample, general equations holding for all frequencies and excitation intensities may be obtained:

$$A=k/[(k+L)^2+\omega^2]^{\frac{1}{2}} \quad \text{and} \quad \tan\phi=\omega/(k+L). \tag{13}$$

If it is assumed that $L=bi$, a plot of $\tan\phi$ as a function of frequency for two different values of current, i, will enable the determination of both k and b. Or if it is possible to measure both the alternating excitation amplitude, ϵ, and the alternating emission intensity amplitude ΔI with sufficient accuracy, it is possible to determine both k and L from one measurement:

$$\frac{\Delta I/I}{\epsilon}=(k/\omega)\sin\phi. \tag{14}$$

Instead of applying the alternating perturbation in the form of an increment of alternating current superposed on a steady current of the exciting beam, it is also possible to vary the steady voltage by an alternating increment. In this case, the equations comparable to Eqs. (3) and (7) are, respectively,

$$L(x,t)=L^*(x)[1+\epsilon(x/KV)\sin\omega t] \tag{15}$$

$$A_0{}^*=[k/(k^2+\omega^2)^{\frac{1}{2}}][1+d/KV] \quad \text{and} \quad \tan\phi_0{}^*=\omega/k. \tag{16}$$

If values of $K=10^{-9}$ cm/volt and $d=10^{-6}$ cm are assumed for Zn_2SiO_4:Mn, after the data of Dowling and Sewell, it is found for a low, steady-excitation intensity that the amplitude of the alternating emission intensity for an alternating perturbation in beam voltage is just twice the amplitude of the alternating emission intensity for an equal alternating perturbation in beam current; the phase shift is the same in both cases.

The main virtue of this type of measurement for the determination of characteristic luminescence quantities, as compared to a steady excitation method, is that in many cases the characteristic quantities can be determined from the measurement of only a phase shift or an incremental intensity rather than from the measurement of total emission intensity.

* Deceased July 10, 1954.

[1] P. H. Dowling and J. R. Sewell, J. Electrochem. Soc. **100**, 22 (1953).

Ion Bombardment-Cleaning of Germanium* and Titanium† as Determined by Low-Energy Electron Diffraction

H. E. Farnsworth, R. E. Schlier, T. H. George, and R. M. Burger
Barus Research Laboratory of Physics, Brown University, Providence, Rhode Island

(Received November 22, 1954)

In the process of carrying out adsorption experiments using germanium and titanium crystals by means of low-energy electron diffraction experiments, it has, of course, been necessary to obtain clean surfaces in high vacuum. After many attempts it became apparent, with both of these crystals, that the usual method of outgassing at elevated temperatures for long periods of time was not sufficient to produce clean surfaces which contain considerably less than one monolayer of gas, as in the case for many metals.

With the titanium crystal, a complicated, hexagonal, single crystalline surface structure was observed after heating at 700° to 750°C for many hours which was oriented on the (0001) hexagonal titanium face. In an attempt to remove this structure, the crystal surface was bombarded by argon ions in a low pressure discharge. After bombardment of a few minutes the diffraction pattern was

nearly obliterated. Subsequently, a short annealing of a few minutes at 500°C resulted in a new diffraction pattern whose surface spacing was the same as that of titanium.[1] Later experiments showed that this is probably a titanium surface which can be maintained for a few hours at sufficiently low pressures. It was further established, by means of the low-energy electron diffraction technique,[2] that the new surface was etched parallel to the (0001) plane, i.e., the one which was parallel to the geometrical boundary, and the same as that which was present originally. This, of course, is one of the most significant aspects of this method of cleaning, where one wishes to expose a desired set of crystal facets. However, it should be emphasized that these results are dependent on the discharge conditions. One is not justified in assuming that the same discharge conditions will produce equally desirable results with surfaces parallel to other crystal planes, until tests have been made by the low-energy electron diffraction method. The annealing conditions are also critical and depend on the mass of the crystal and its mounting.

It was observed in the above experiments that during the short annealing subsequent to bombardment, the residual pressure in the tube was increased to a considerably higher value than that produced by a similar annealing not preceded by bombardment. This is interpreted as an indication that argon penetrates the crystal lattice during bombardment.

In order to obtain a clean surface of titanium it was necessary to precede the ion bombardment by a thorough high temperature outgassing to remove absorbed gas. When this was not done the diffraction pattern obtained after bombardment and annealing was inferior.

In the experiments on the (001) face of a germanium crystal, it was observed that a weak diffraction pattern could be obtained after outgassing the crystal for several hours at 600° to 700°C. This diffraction pattern appeared to be characteristic of the germanium lattice. It was never possible to increase the intensities of the small number of observed beams to the point where the identification could be made with assurance. In fact, in many cases of prolonged heating near the melting point, the pattern became weaker or disappeared entirely. It was, therefore, concluded that the surface structure was not that of clean germanium but of a compound with a relatively imperfect crystal structure resembling germanium to some extent.

Subsequently, the aforementioned method of cleaning by ion bombardment and annealing was attempted for the germanium crystal. In the first attempt hydrogen ions were used instead of argon, since the experimental tube was equipped with a palladium tube. The results were negative. A gas-handling system was then installed so that argon could be admitted under controlled conditions. The results obtained after argon ion bombardment are very similar in nature to those obtained with titanium. The contaminating surface layer was removed by the ion bombardment, but the diffraction pattern was obliterated. After a subsequent short annealing at 500°C, a sharp intense diffraction pattern was observed which we believe to be characteristic of a clean germanium crystal surface. This surface can also be kept stable under best vacuum conditions for periods of several hours before adsorption effects from the residual gases become observable.

The structures of these adsorbed gases are now being investigated for both germanium and titanium with the use of gas handling systems for known gases.[3]

* This part of the research was supported by a Joint Service's Contract with the Massachusetts Institute of Technology and a subcontract with Brown University.

† This part of the research was supported by Office of Ordnance Research, U. S. Army.

[1] This observation was first made with a diffraction unit operated by R. E. Schlier.

[2] H. E. Farnsworth, Phys. Rev. **49**, 604 (1936).

[3] We are aware of the fact that ion bombardment has been used previously to clean surfaces, but to our knowledge the method of low-energy electron diffraction has not been used to determine the condition of the surface after low-temperature annealing. For information on penetration of low-energy diffracted electrons and gas adsorption on copper and nickel crystals, see the following. H. E. Farnsworth Phys. Rev. **49**, 605 (1936); **35**, 1133 (1930); R. E. Schlier and H. E. Farnsworth, J. Appl. Phys. **25**, 1333 (1954); two papers in Bull. Am. Phys. Soc. **29**, 35 (1954).

Errata: Relativistic Dynamics of a Charged Particle in Crossed Magnetic and Electric Fields with Application to the Planar Magnetron

[J. Appl. Phys. 25, 683, 691 (1954)]

Louis Gold
Massachusetts Institute of Technology, Cambridge, Massachusetts

FOLLOWING notification by Dr. J. A. Bradshaw and Dr. J. F. Hook of several errors appearing in two companion papers, the author scrutinized the texts to eliminate other possible errors that might inconvenience the reader. The findings are summarized below:

First paper:

1. Equation (35) should have the first square bracket $[\ ]^{\frac{1}{2}}$—exponent $\frac{1}{2}$ omitted in paper.
2. Equation (44) should have term in denominator read $c/\omega L$ rather than c/ω^2.
3. Equation (46b) should read

$$\frac{(1+\gamma^2)^{\frac{1}{2}}}{\omega}[\ ]$$—exponent $\frac{1}{2}$ omitted in paper.

4. Equation (49) should read†‡

$$y=\frac{c\gamma}{\omega(1-\gamma^2)^{\frac{1}{2}}}\left[\frac{(1-\gamma^4)^{\frac{1}{2}}\cos\phi}{1+\gamma^2\sin\phi}+2\left\{\tan^{-1}\left[\frac{\tan\frac{\phi}{2}+\gamma^2}{(1-\gamma^4)^{\frac{1}{2}}}\right]-\tan^{-1}\left(\frac{1+\gamma^2}{1-\gamma^2}\right)^{\frac{1}{2}}\right\}\right].$$

5. Equation (50) should read†‡

$$y^*=\frac{\pi\gamma c}{\omega(1-\gamma^2)^{\frac{1}{2}}}.$$

6. Equation (51) should read†‡

$$\frac{x^*}{y^*}=\frac{2}{\pi}(1-\gamma^2)^{\frac{1}{2}}.$$

7. Four lines below Eq. (51) should read,†‡ "ever-decreasing spikes so that $x^*/y^*\to 0$."
8. Two lines below Eq. (68) *Lh* should read *LH*.
9. These corrections should carry over to the summary section.

Second paper:

1. The right-hand member of Eq. (6) should read $(C_0/2)\left(\frac{r^2}{2}c_0+r^2-1\right)$.
2. Equation (7) should read $\epsilon=(1+B)^{\frac{1}{2}}$.
3. Relation (8) should read $C_0\leq 0$.
4. On page 692, line 2 should have *semi*-axes with

$$\left[\frac{C_0}{2}\left(\frac{r^2}{2}C_0+r^2-1\right)\right]^{\frac{1}{2}}\text{ minor}$$

$$\left[-\left(\frac{r^2}{2}C_0+r^2-1\right)\right]^{\frac{1}{2}}\text{ major.}$$

5. Equation (12) should have second bracket read

$$\frac{C_0}{2}r-\left[\frac{C_0}{2}\left(\frac{r^2}{2}C_0+r^2-1\right)\right]^{\frac{1}{2}}\sin\theta.$$

6. Equation (19) should have factor 2 before $\beta_{0y}{}^2$ removed.
7. Equation (20) in the denominator of ( ) should have $[1-(r/2)\beta_{0y}]$.

8. In (b), C_0 value should have no 2 before β_0^2.
9. In (d), C_0 value should have factor $\frac{1}{2}$ before β_0^2 in the numerator.
10. In (d), θ_0 rel'n should have ( ) in denominator read $[1-(\sqrt{2}/4)r\beta_0]$.
11. Equation (25) should read‡

$$y=\frac{2\sqrt{2}}{|C_0|\,(C_0+2)^{\frac{1}{2}}}\cdot\frac{c}{\omega}\cdot\frac{1}{1-\delta^2}\left[\delta A+\frac{2+C_0(1-\delta^2)}{(1-\delta^2)^{\frac{1}{2}}}B\right]$$

where

$$A=\frac{\cos\theta}{1+\delta\sin\theta}-\frac{\cos\theta_0}{1+\delta\sin\theta_0},$$

$$B=\tan^{-1}\left[\frac{\tan\frac{\theta}{2}+\delta}{(1-\delta^2)^{\frac{1}{2}}}\right]-\tan^{-1}\left[\frac{\tan\frac{\theta_0}{2}+\delta}{(1-\delta^2)^{\frac{1}{2}}}\right].$$

12. In expression (29) y^* should read‡

$$y^*=\frac{2\sqrt{2}}{|C_0|\,(C_0+2)^{\frac{1}{2}}}\cdot\frac{c}{\omega}\cdot\frac{1}{1-\delta^2}\left[\frac{-\delta\cos\theta_0}{(1-\delta\sin\theta_0)}+\frac{2+C_0(1-\delta^2)}{(1-\delta^2)^{\frac{1}{2}}}\left\{\tan^{-1}\left(\frac{\delta-1}{(1-\delta^2)^{\frac{1}{2}}}\right)-\tan^{-1}\left[\frac{\tan\frac{\theta_0}{2}+\delta}{(1-\delta^2)^{\frac{1}{2}}}\right]\right\}\right].$$

13. In integral (38) the dx should be outside the bracket.
14. In integral (40) the factor $(1-x)^2$ in the denominator should not appear.
15. Equation (65) should read $\left|\int_{L_1}^{L_2}\frac{2}{\gamma^3(1-\gamma^2)}\left(\frac{1-\gamma^2-\gamma^2Z}{Z}\right)\cdots\right.$ etc.
16. The first term of Eq. (68) should read

$$\frac{2[3\gamma+2(1+\gamma^2)^{\frac{1}{2}}]}{(1-\gamma^2)^2(1+3\gamma^2+2\gamma^4+2\gamma(1+\gamma^2)^{\frac{1}{2}})}.$$

The author is most appreciative of the interest shown by Dr. J. A. Bradshaw and Dr. J. F. Hook in bringing to light the misrepresentations indicated. He also wishes to thank Virginia Johnson and Janet Beck of the Lincoln Laboratory, M.I.T. for their painstaking efforts to ferret out other errors.

† Called to my attention by Dr. J. F. Hook, Hughes Aircraft, Culver City, California.
‡ Called to my attention by Dr. J. A. Bradshaw, 2221 Stone Ridge Road, Schenectady, New York.

Books Reviewed

The Organization of Applied Research in Europe, The United States, and Canada. Organization for European Economic Cooperation, Paris, 1954 (order from O.E.E.C. Mission Publications Office, 2002 P Street N.W., Washington 6, D.C.)
Vol. I. A Comparative Study, Pp. 76. Price $1.00.
Vol. II. Applied Research in Europe, Pp. 181. Price $2.00.
Vol. III. Applied Research in the United States and Canada. Pp. 125. Price $1.25.

These studies are reports of technical assistance missions of the O.E.E.C. Volume III was prepared by a committee of the National Research Council. Expenditures are tabulated in the various countries for research by government agencies and by industries and universities. Typical organization of research in large and small companies and in government laboratories are described.

Flash! Seeing the Unseen by Ultra High-Speed Photography. Harold E. Edgerton and James R. Killian, Jr. Pp. 215, Second Edition. C. T. Branford Company, Boston, 1954. Price $6.50.

In the fourteen years since the first edition of this book, electronic flash photography has become a popular and important technique in scientific and industrial photography. This book contains hundreds of photographs, some of which are in color, illustrating applications of microsecond flash techniques to all fields of photography. Most of the pictures show how this technique permits the analysis of extremely rapid motion, such as in the flight of a humming bird or the break-up of a jet of water. A supplementary section at the end of the book describes the elementary theory of electronic flash photography and describes equipment in current use.

Optics. A. Sommerfeld. Pp. 383+xii. Academic Press, New York, 1954. Price $6.80.

This text is Volume IV in the six volume series of *Lectures on Theoretical Physics* by the author. It is a translation of the lectures given by Professor Sommerfeld at the University of Munich in 1934. The work is closely connected with *Electrodynamics*, Volume III, and the formalism and intrinsic character of Maxwell's equations treated in that volume are assumed to be known.

Reflection and refraction of light are discussed first. Fresnel's formulas, in this connection, are examined in detail. Also included in this chapter are sections on standing light waves and colors of thin membranes and thick plates.

A chapter dealing with the theory of dispersion has, in addition to the classical treatment, a wave-mechanical description of the phenomenon. Optics of moving media and sources (together with astronomical topics) and crystal optics constitute two other chapters.

A lengthy discussion of the theory of diffraction is presented in the last part of the book. The topics of Čerenkov radiation and geometrical optics are also discussed.

Problems and their solutions are given.

Symposium on Fluorescent X-Ray Spectrographic Analysis. ASTM Special Technical Publication No. 157. Pp. 72. American Society for Testing Materials, Philadelphia, 1954. Price $1.75.

These papers were presented at the 56th Annual Meeting of the American Society for testing Materials held in Atlantic City, New Jersey, June 1953, in the symposium sponsored by ASTM Committee E-2 on emission spectroscopy. Various aspects of the basic theory and fundamentals of fluorescent x-ray spectrographic analysis and its applications are discussed. Pictures, graphs, charts, tables, and schematic diagrams are included.

Graphs for Calculations of Compressible Airflow. L. Rosenhead. Pp. 115+x. (11½ in.×15 in.). Oxford University Press, New York, 1954. Price $13.45.

This is a companion volume to *A Selection of Tables for Use in Calculations of Compressible Airflow*. The object of both books is to furnish data in easily used form for research and design involving the flow of air in which compressibility effects are important. Large, easily read graphs are presented for isentropic flow, normal shocks, oblique shocks, conical flow, and Reynolds numbers. Some of the graphs are on a single page and are intended to give a quick appreciation of the variations of the quantities and to present data for rough calculations. Other graphs are multiple-page plots and present the data with an accuracy approaching that of the volume of *Tables*. This book was prepared on behalf of the Aeronautical Research Council by the Compressible Flow Tables Panel, L. Rosenhead, *chairman*.

The Psychology of Invention in the Mathematical Field. JACQUES HADAMARD. Pp. 145+xiii. Reprint of first edition, Dover Publications, Inc. (first edition, copyright 1945, by Princeton University Press). Price $1.25, paperbound; $2.50, clothbound.

This book is concerned with the origin and fruition of important ideas, especially in mathematics. It was inspired by a lecture on the subject by Henri Poincare. Frequent references are made to Poincare's exposition of his own mental processes, and reports by Pascal, Dalton, and many others are included. One of the appendices is a letter from Albert Einstein in response to a questionnaire printed in a French journal of mathematics in 1902 and 1904. The questionnaire is given as Appendix I and numbers thirty questions addressed to mathematicians about their training, goals, work habits, and additional interests. The third appendix gives a classic example (the infinitesimal calculus) of a feature of invention treated at several points in the main portion of the essay: the failure of the discoverer to perceive the generality and applicability of what he has found.

Die Grundlagen der Akustik. EUGEN SKUDRZYK. Pp. 1084 +xxii. Springer-Verlag, Vienna, 1954. Price $35.00.

This is a comprehensive reference book and treats in encyclopedic fashion the whole field of acoustics. The references alone occupy 125 pages. The usual topics in acoustics are considered and also some relatively unusual ones. Internal friction, sound absorption in water, and strongly nonlinear behavior are some examples. The theory of acoustics is stressed throughout the book, and very little attention is paid to apparatus and measurements.

Sensations of Tone. H. HELMHOLTZ. Pp. 576+xix, 68 diagrams. Dover Publications, Inc., 1954. Price $4.95.

The present volume is a reprint of the second English edition of 1885 which was translated by A. J. Ellis from the fourth German edition. This now classic treatise on tone brings together a command of several different fields of study necessary for a comprehensive understanding of what we hear and why we hear it. Helmholtz, though both a physiologist and a physicist, was in addition a student of music and has drawn from all three areas to present a unified description of sound and its perception as music by the ear. The book describes many of Helmholtz's own experiments on hearing and his conclusions concerning discord and harmony, and his concept of esthetics. The translator has profusely footnoted the volume and added a large appendix, primarily for the benefit of students of music, toward whom about one half of the book is directed.

Physical Methods of Organic Chemistry. Part III

EDITED BY ARNOLD WEISSBERGER. Pp. 434+xi, second edition. Interscience Publishers, Inc., New York, 1954. Price $8.50.

This book is the third part of the first volume in a series entitled *Technique of Organic Chemistry*. This series is devoted to a comprehensive presentation of the techniques which are available for the investigation of organic compounds. The object of the contributors to Volume I has been to provide a description of tested physical methods, and the information necessary for a critical evaluation of the experimental results. Part III contains additions to the material in Parts I and II as well as fields not previously treated.

Chapters are included on electron microscopy, microspectroscopy, radio-frequency spectroscopy, neutron diffraction, and measurements of dielectric constants and loss. Supplementary material is given on x-ray and electron diffraction techniques, viscometry, radioactivity, and magnetic susceptibility measurements.

General Theory of High Speed Aerodynamics

EDITED BY W. R. SEARS. Pp. 758+xiv. Princeton University Press, Princeton, 1954. Price $15.00.

This is Volume VI of the series *High Speed Aerodynamics and Jet Propulsion*. This volume summarizes present-day knowledge in the theory of high speed aerodynamics. It begins with an overall sketch of the subject by Theodore von Kármán. Largely nonmathematical, Professor von Kármán's section touches on the major features of subsonic, transonic, supersonic, and hypersonic flows. It emphasizes important recent developments as well as some problems still unsolved. K. O. Friedrichs presents a new study of the mathematical nature of hyperbolic flows, that is, supersonic or unsteady compressible flows.

Following these introductory sections, attention is directed especially toward the theories based on the small disturbance hypothesis. W. R. Sears presents and summarizes subsonic small perturbation theory, with emphasis on recent studies that unify and clarify this subject, and on rotational and unsteady flows. The largest section of the volume is by Max. A. Heaslet and Harvard Lomax, who present, in a unified manner, the whole subject of linearized supersonic theory, especially with application to wings. Their section also includes the transonic and hypersonic nonlinear approximations. M. J. Lighthill's contribution concerns the methods of higher approximation which are used to improve the first order theories and to explain and eliminate their shortcomings.

Departing now from the ideas of perturbation theories, Y. H. Kuo and W. R. Sears collaborate in a section presenting a review of a highly complicated and confused subject: the theory of potential subsonic and mixed supersonic-subsonic plane flows based on the hodograph method. Finally, in two sections, Antonio Ferri applies the method of characteristics to steady two- and three-dimensional flows and flow patterns characterized by the presence of shock waves. These sections include Professor Ferri's recent work on the linearized methods of characteristics.

Textbook of Physics

EDITED BY R. KRONIG. Pp. 855+xiii. Interscience Publishers, Inc., New York, and Pergamon Press Ltd., London, 1954. Price $10.00.

This volume is an English translation of the Dutch language textbook which surveys physics as a whole at the underclass, undergraduate level. The contributors are J. De Boer, H. C. Burger, P. H. Van Cittert, C. J. Gorter, A. C. S. Van Heel, R. Kronig, P. Van Der Leeden, G. J. Sizoo, and J. Korringa.

The arrangement is to first treat the classical phenomenological parts of physics, such as, mechanics, theory of vibrations, electrodynamics, and optics. These are followed by chapters on theory of atomic structure, atomic theory of heat, atomic electricity, and thermodynamics. Sections are included on electrical and optical instrumentation. The final chapter surveys the field of medical physics.

The rationalized Giorgi system based on four fundamental units is used throughout the book.

Index to the Literature on Spectrochemical Analysis, Part III, 1946–1950. B. F. SCRIBNER AND W. F. MEGGERS. ASTM Special Technical Publication No. 41-C, Pp. 226+iii. American Society for Testing Materials, Philadelphia, 1954. Price $4.50.

This is the third part of a series of bibliographical surveys of the literature of spectrochemical analysis covering the period 1946–1950. References and abstracts are listed chronologically by year and in alphabetical order of the first author's name in each year. A total of 1264 references are included. Both subject and author indexes are provided.

Books Received

The Nation Looks at Its Resources. Pp. 418+xii. Resources for the Future, Inc., Washington, D. C., 1954. Price $5.00.

Fluid Dynamic of Jets. SHIH-I PAI. Pp. 227+xv, with Figs. D. Van Nostrand Company, Inc., New York, 1954. Price $5.00.

Physical Methods and Organic Chemistry. ARNOLD WEISSBERGER. **Technique of Organic Chemistry, Vol. 1, Part III.** Pp. 2097–2530+xi, numerous Figs., second edition. Interscience Publishers, Inc., New York, 1954. Price $8.50.

Rocket Exploration of the Upper Atmosphere. R. L. F. BOYD and M. J. SEATON. Pp. 376+viii, numerous Figs. Interscience Publishers, Inc., New York, 1954. Price $11.00.

Practical Photography. B. K. JOHNSON. Pp. 146, Figs. 41, Longmans Green and Company, New York, 1954. Price $1.80 (text edition), $2.40 (trade edition).

Sailing Aerodynamics. JOHN MORWOOD. Pp. 124, Fig. 82. Philosophical Library, New York, 1954. Price $7.50.

Grundzuge der Zerspanungslehre. MAX KRONENBERG. Pp. 430+xxvi, Figs. 293, second edition. Springer-Verlag, West Berlin, 1954.

Announcements

Annual Winter Conference

The **Annual Winter Conference of the New York Section, ISA,** will be held on Wednesday, February 23, 1955, at the Hotel Statler in New York City. The theme of the Conference is, "Instrumentation for Data Reduction and Presentation." Please make inquiries to Mr. Hamilton Bristol, General Chairman, c/o Bristol Company, 250 West 57th Street, New York City.

Western Computer Conference

The **Joint Western Computer Conference and Exhibit** will be held at the Statler Hotel, Los Angeles, California, on March 1, 2, and 3, 1955, sponsored by the I.R.E., A.I.E.E., and Association for Computing Machinery. Those interested in attending should write: William Gunning, *Conference Secretary*, International Telemetering Corp., 2000 Stoner Avenue, Los Angeles 25, California.

The **American Association of Spectrographers** is planning their 6th Annual Conference in Chicago, Illinois, May 6, 1955, on the subject "Industrial Applications of Spectroscopy." Contributed papers in the fields of emission, x-ray fluorescence, or absorption spectroscopy, as applied to industry, are invited. Abstracts must be submitted by March 1, 1955. Please address all inquiries to:

F. E. Stedman or E. E. Stilson, Co-Chairman
Engineering Research Laboratory
Bendix Products Division
Bendix Aviation Corporation
401 North Bendix Drive
South Bend 20, Indiana

Government-Industry Research Committee

The National Academy of Sciences-National Research Council has been asked by officials of both industry and government to establish within the Academy-Research Council framework, a committee to concern itself with the common interests and relationships of industrial and governmental research, particularly in the area of applied research.

Conferences between industrial and governmental research executives and directors, called by the Academy-Research Council, recommended that a small committee be organized to explore the need for better acquaintance and understanding between Government and industry research leaders, and to consider methods for accomplishing this objective. As a result, the Government-Industry Research Committee has been organized by the Academy-Research Council with the following membership:

Edgar C. Bain, United States Steel Corporation, Chairman
Allen V. Astin, National Bureau of Standards
D. P. Barnard, Deputy Assistant Secretary of Defense, Research and Development
Ralph Bown, Bell Telephone Laboratories, Inc.
Ralph Connor, Rohm and Haas Company
Hugh L. Dryden, National Advisory Committee for Aeronautics
Paul D. Foote, Gulf Research and Development Company
G. E. Hilbert, Agricultural Research Service, U. S. Department of Agriculture
Randolph Major, Merck and Company, Inc.
Roy C. Newton, Swift and Company
Alan T. Waterman, National Science Foundation

At its first meeting the Committee concluded that effective mechanisms already exist in many fields for furthering mutually helpful relations between government and industry research. However, the Committee agreed to hold itself available as necessary to assist in exchanging views and ideas designed to improve such relations where either Government or industry groups may feel this to be desirable.

When its services are requested, the Committee proposes to consider first the extent to which the need can be satisfied by existing mechanisms. If appropriate, the Committee will then consider designation of an *ad hoc* group of individuals active in the particular field concerned to assist in bringing about improved understanding and closer relationships between Government and industry people in that field.

Cover Photograph

The photograph shows an experimental apparatus now in use at Bell Telephone Laboratories in a program of study of the electron ejection from metal surfaces by slowly moving positive ions. In this device ions formed by electron impact in a gas are accelerated and focused by suitable electrostatic lenses; they strike a target which is situated inside the spherical electron collector to be seen in the right foreground inside the glass envelope. No m/e analysis of the ion beam is provided, but a pure beam of singly charged ions of the noble gases may be obtained by using bombarding electron energies below the second ionization energy of the gas. Total yield of ejected electrons per incident ion and energy distributions of these electrons may be measured at any incident ion energy from 10 to 1000 electron volts. Constructional features of interest are the minimum use of the glass envelope as a structural element, all electrodes except the source filament and target being mounted from the metal center section, and the mounting of electrodes on quartz rods which provide accurate alignment of slits and apertures. Differential pumping is provided between the target chamber (right) and the acceleration chamber (left). Gas is admitted directly into the ionization chamber through the tube which involves the metal bellows at the left. With careful processing a measured pressure of background gases of 1×10^{-10} mm Hg has been achieved on the pumps. The monolayer adsorption time for adsorbable gases is then greater than 10 hours (note that at 10^{-6} mm Hg this time is 1 second).

The foregoing information has been supplied by Dr. H. D. Hagstrum of Bell Telephone Laboratories.